When AI Improves Answers but Slows Knowledge Creation: A Decomposition of AI's Impact on Digital Public Goods

Keh-Kuan Sun [*†]

March 25, 2026

Abstract

Generative AI helps users solve problems more efficiently, but without leaving a public trace. Fewer discussions and solutions reach public platforms, and the archives that future problem-solvers depend on can shrink. We build a dynamic model of public good provision where agents contribute by solving problems that other agents posted on a public platform, and the accumulated solutions form a depreciating public archive. AI reduces archive creation through two margins that require different instruments. The flow margin: the posted volume of knowledge-enhancing queries declines as AI resolves more problems privately before they reach the platform. The resolution margin: the probability that posted queries are resolved declines as AI raises contributors' outside options, thinning the contributor pool and creating congestion on the platform. The two margins interact through a self-undermining feedback that can generate low-archive traps. The decomposition yields a diagnostic prediction: in the congested regime, a joint decline in posted volume and conditional resolution requires that supply-side pool thinning is quantitatively present, whereas volume decline with stable or rising resolution indicates that private diversion alone is the dominant force. Encouraging public sharing of AI-assisted solutions offsets the decline associated with private diversion but cannot repair participation-driven deterioration in conditional resolution, which requires maintaining contributor engagement directly.

Keywords: generative AI; knowledge commons; digital public goods; dynamic crowd-out; congestion matching; public logging.

JEL: D62; D83; H41; O33.

[*]Department of Economics, Fairfield University. Email: ksun@fairfield.edu

[†]The author has no sources of research support, financial relationships, or potential conflicts of interest to disclose.

1 Introduction

Generative AI helps users solve problems more efficiently, but without leaving a public trace. The immediate benefit is well documented: faster resolution, lower search and verification costs, and measurable productivity gains in knowledge work (Brynjolfsson, Li, and Raymond 2025; Dillon et al. 2025). But when more problems are solved privately, fewer problem-solving episodes are converted into reusable public knowledge for others. Consider a student who raises a question in a lecture hall: the instructor answers, and every other student in the room learns something. Or consider a new hire who brings a recurring problem to a manager; after fielding the same question several times, the manager writes it up as an internal FAQ, and future employees never need to ask. In both cases, the act of asking and answering in a shared setting produces a durable record that benefits people who never asked. Open knowledge platforms operate on the same principle: on Stack Overflow, for example, a posted question and its accepted answer become a public artifact that thousands of future visitors can reuse. When AI resolves that same question in a private chat, the answer may be just as good for the person who asked, but no public artifact is created, no future visitor benefits, and the platform's archive grows a little more slowly. Scale this up across millions of queries per day, and the cumulative effect on the stock of open, platform-mediated knowledge can be substantial.

We build a dynamic model of public good provision where agents contribute by solving problems that other agents posted on a public platform, and the accumulated solutions form a depreciating public archive. AI reduces archive creation through two margins that require different instruments. The *flow margin* is the decline in the posted volume of knowledge-enhancing queries that reach the platform; the primary mechanism is private diversion, in which AI resolves problems that would otherwise have been posted publicly. The *resolution margin* is the decline in the probability that posted questions get answered; the primary mechanism is contributor pool thinning, in which AI raises the return to working alone, contributors exit the platform, and the resulting congestion leaves posted questions less likely to encounter a capable answerer. Because posted flow and resolution probability are equilibrium objects, the two margins are not causally independent: private diversion also eases congestion (pushing resolution upward), and pool thinning can feed back into posting incentives. A self-undermining feedback amplifies the shrinkage: as the archive grows, AI tools that draw on it improve, diverting more questions from posting and raising the opportunity cost of contributing further. The two margins require

different instruments: encouraging public sharing of AI-assisted solutions offsets the decline associated with private diversion by converting off-platform resolutions into public artifacts, but it does not repair congestion- or participation-driven deterioration in conditional resolution, which requires maintaining contributor engagement on the platform.

Online Q&A platforms offer a clean setting for studying these two margins because every interaction either produces a public artifact or does not. Evidence that AI is active on both margins is accumulating across platforms. Del Rio-Chanona, Laurentsyeva, and Wachs (2024) document a sharp decline in Stack Overflow activity relative to less exposed Q&A platforms after the release of ChatGPT. Padilla et al. (2025) document a 20% or larger drop in search traffic to knowledge sites as users divert queries to LLMs. Lyu et al. (2025) show that Wikipedia articles whose content overlaps with ChatGPT's capabilities experienced significant declines in viewership and editing activity relative to dissimilar articles. The breadth of this evidence motivates a model that isolates the mechanism at a level of generality beyond any single platform.

The model distinguishes between routine questions, which are standard and frequently asked, and knowledge-enhancing questions, whose resolution generates reusable public knowledge. AI improves private problem-solving for both types of questions, but resolves routine questions at a higher rate. This has two simultaneous effects: it frees agents from routine tasks, raising the opportunity cost of time they would otherwise spend contributing on the platform, and it changes the composition of questions that reach the platform, since routine questions are disproportionately diverted. Agents who fail to solve a problem privately decide whether to post it, trading off a posting cost against the expected value of a public answer. On the platform, posted questions are matched to contributors under a congestion technology: each contributor handles at most one question per period, so when contributors exit, each remaining question is less likely to encounter an answerer. A single domain parameter, the routine-task share π, governs both how many questions AI diverts from posting and how much AI raises the return to working alone, generating a cross-platform prediction about which types of platforms are most exposed.

The paper's main results organize around the two margins. Lemma 1 builds on J. S. Gans (2024), who argues that when AI diverts easy questions to private resolution, the remaining posted pool shifts toward harder problems and contributors become more willing to answer. In our dynamic setting, the necessary and sufficient condition for this selection logic to raise

the answering cutoff is that the posted pool becomes richer in knowledge-enhancing questions. Proposition 1 shows that when participation is endogenous, congestion from contributor pool thinning can overwhelm this composition improvement, so the sign of the resolution change is the outcome of a race among three forces: composition enrichment (raises resolution), posted-flow reduction (raises resolution by easing congestion), and pool thinning (lowers resolution). Proposition 2 and Corollary 1 establish dynamic crowd-out and decompose the decline in archive creation into the flow margin and the resolution margin. Proposition 3 shows that a self-undermining feedback (a richer archive raises outside options, thinning the contributor pool) can generate multiple steady states, including a structural minimum viable archive below which the platform cannot sustain itself. This threshold exists in both economies; AI's effect is to compress the stable steady state inward, narrowing the viable region. Proposition 4 shows that encouraging public sharing of AI-assisted solutions addresses the flow margin but not the resolution margin, and identifies a threshold above which the structural minimum viable archive is eliminated and the viable region expands. Across platforms, the routine-task share π scales both margins simultaneously, making high-π platforms doubly vulnerable.

A growing literature studies how AI reshapes the economics of knowledge production. Abis and Veldkamp (2023) argue that AI changes the knowledge production function in a manner analogous to industrialization: just as mechanization raised the returns to capital relative to labor, AI raises the returns to data relative to human judgment, transforming who produces knowledge and how. The present paper studies a downstream consequence: when AI shifts knowledge production toward private channels, the public archives that future problem-solvers depend on can shrink. More recently, Acemoglu, Kong, and Ozdaglar (2026) develop a dynamic model in which AI substitutes for human learning effort, generating knowledge collapse when private problem-solving displaces the public signal that sustains general knowledge. The instrument that emerges from their analysis is garbling (deliberately degrading AI accuracy to incentivize human effort). Our two-margin separation leads to a different instrument: encouraging public sharing of AI-assisted solutions, which preserves full AI capability while converting private resolutions into public artifacts. The frameworks are complementary: they provide general-equilibrium welfare analysis that the present partial-equilibrium model cannot, while the two-margin decomposition here identifies platform-level mechanisms that their aggregate framework does not separate. We discuss a more detailed comparison in Section 6.

The rest of the paper is organized as follows. Section 2 introduces the model. Section 3

derives the two margins: the composition effect from selective diversion (Lemma 1), the race that determines the sign of the resolution change (Proposition 1), dynamic crowd-out (Proposition 2), and the decomposition into the flow and resolution margins (Corollary 1). Section 4 studies the self-undermining feedback and establishes the multiple-steady-state result (Proposition 3). Section 5 analyzes public sharing of AI-assisted solutions and its limits (Proposition 4). Section 6 discusses scope, cross-channel generalization, and empirical diagnostics.

2 Model

Time is discrete, $t = 0, 1, 2, \ldots$. The platform maintains a public archive stock $K_t \geq 0$, which depreciates at rate $\lambda \in (0, 1)$ per period and lowers the cost of producing public answers. Depreciation is interpreted as *effective archive depreciation*: obsolescence (APIs and libraries update), link rot, outdated solutions, moderation and closure, and decay in search visibility that reduces reuse of old answers. The core dynamic object is $h(K) - \lambda K$, where $h(K)$ is expected new public knowledge created at state K; steady states satisfy $h(K) = \lambda K$ and provide a "platform viability" benchmark.

2.1 Domain and agents

The platform serves a knowledge domain characterized by a routine-task share $\pi \in (0, 1)$. A fraction π of problems in the domain are routine (type L): standard lookups, well-documented procedures, frequently asked questions. For instance, looking up the syntax for a specific function in a difference-in-differences package, or finding the correct option to cluster standard errors in a regression command. The remaining fraction $1 - \pi$ are knowledge-enhancing (type H): novel problems whose resolution generates reusable public knowledge. For instance, working out a credible identification strategy for a new empirical setting, or diagnosing why a particular estimator behaves unexpectedly when treatment timing is staggered. The parameter π is a property of the domain (e.g., introductory programming vs. research mathematics), not of individual agents.[1]

1. In domains where users and contributors are drawn from distinct populations (e.g., students asking and professors answering), the routine-task share relevant to query composition and the one relevant to contributors' outside options may differ. This can be accommodated by introducing a separate π_c for the contributor side. The baseline assumes a self-serving community where agents are simultaneously users and potential contributors,

There is a unit mass of agents, each indexed by persistent ability $\alpha_i \sim \Psi$ with density $\psi > 0$ on $[\underline{\alpha}, \bar{\alpha}]$. Each period, each agent faces a task drawn from the domain: type L with probability π, type H with probability $1-\pi$. This generates a unit mass of queries. Within type $\theta \in \{L, H\}$, a query draws residual difficulty $r \sim G_\theta$ with density $g_\theta > 0$ on $[r_L, r_H]$. Residual difficulty governs amenability to private resolution (search engines, AI, documentation); it does not determine the cost of answering on the public platform, which depends on the idiosyncratic match quality between question and answerer (Section 2.4). Each agent is also a potential contributor: after facing their own task, they may be matched to someone else's posted query on the platform.

The routine-task share π enters the model through two channels. First, it determines query composition: fraction π of arriving queries are type L (which do not expand the archive) and fraction $1-\pi$ are type H (which do). Second, it governs the magnitude of the outside-option shift when AI arrives: in high-π domains, AI automates a larger share of agents' own workloads, freeing more time for own knowledge-enhancing work and raising the opportunity cost of platform volunteering. Both channels imply that platforms in high-π domains are more vulnerable to AI-induced archive decline.

2.2 Outside options and participation

Each agent i has a per-period outside-option payoff $w^e(\alpha_i, K; \pi)$, representing the total return from devoting the period to own work instead of platform volunteering. The function w^e depends on ability α_i, the archive stock K, the domain's routine-task share π, and the environment $e \in \{\text{HO}, \text{AI}\}$. It is continuously differentiable with four maintained properties:

(i) $w^e_\alpha > 0$: higher-ability agents earn more from own work.

(ii) $w^e_K > 0$: a richer public archive improves retrieval-augmented and AI-assisted tools that complement own work.

(iii) $w^{\text{AI}}(\alpha, K; \pi) \geq w^{\text{HO}}(\alpha, K; \pi)$ for all (α, K): AI weakly improves outside options.

(iv) $w^{\text{AI}}(\alpha, K; \pi) - w^{\text{HO}}(\alpha, K; \pi)$ is increasing in π: the AI-induced outside-option shift is larger in high-π domains.

so a single π governs both sides; this is the empirically relevant case for platforms like Stack Overflow and for within-firm knowledge bases where employees both ask and answer.

The shift $w^{\text{AI}} - w^{\text{HO}}$ captures the net effect of AI on the opportunity cost of volunteering; in high-π domains, AI automates a larger share of agents' routine workloads, freeing time that agents redirect to own knowledge-enhancing work. The reduced-form specification encodes this time-reallocation logic in the maintained comparative statics without tracking time budgets explicitly.[2]

Let $S^e(K)$ denote the expected per-period payoff from platform volunteering at archive state K in environment e (derived from the answering game below; this payoff does not depend on α because ability affects the outside option, not on-platform returns). Agent i participates on the platform iff $S^e(K) \geq w^e(\alpha_i, K; \pi)$. Because w^e is increasing in α, there exists a unique participation cutoff $\alpha^{*,e}(K)$ defined by

$$S^e(K) = w^e(\alpha^{*,e}(K), K; \pi). \tag{1}$$

Agents with $\alpha_i \leq \alpha^{*,e}(K)$ participate; those with $\alpha_i > \alpha^{*,e}(K)$ exit to own work. The measure of participating agents is $\Psi(\alpha^{*,e}(K))$.

AI shifts the participation cutoff downward through two reinforcing channels: the direct channel ($w^{\text{AI}} > w^{\text{HO}}$ raises the outside option for all agents, making the participation condition harder to satisfy) and a feedback channel ($w^e_K > 0$ means that a richer archive further raises outside options, amplifying exit pressure).

2.3 Private resolution and endogenous posting

In each environment $e \in \{\text{HO}, \text{AI}\}$, an agent whose task becomes a query first attempts private resolution using an outside option. In the pre-AI benchmark (HO), the outside option captures search engines, documentation, coworkers, and other off-platform resources; in the AI regime (AI), the outside option is stronger because it includes an LLM front end in addition to baseline

2. An explicit continuous time-allocation model would derive the π-scaling structurally and characterize whether the outside-option shift is concave or convex in π, but it would require a continuous time-choice margin (how much time to devote to volunteering vs. own work), which substantially complicates the participation structure. The reduced-form approach delivers the qualitative comparative statics that the cross-platform predictions require; a structural time-allocation extension would be needed to quantify the π-vulnerability prediction rather than merely sign it.

resources. Private resolution succeeds for a type-θ query whenever

$$r \le \bar{r}_\theta^e(K), \tag{2}$$

where $\bar{r}_\theta^e(K)$ is an environment-specific outside-option reliability cutoff that may depend on the archive stock K. We interpret $\bar{r}_\theta^e(K)$ as an *effective reliability* cutoff: higher hallucination or miscalibration risk lowers the effective cutoff, equivalently lowering the user's expected value from relying on private resolution (J. Gans 2026). The probability of private resolution in environment e is $a_\theta^e(K) \equiv G_\theta(\bar{r}_\theta^e(K))$. AI is modeled as a shift in outside-option quality: $a_\theta^{\mathrm{AI}}(K) \ge a_\theta^{\mathrm{HO}}(K)$ for the relevant tasks.

If private resolution fails ($r > \bar{r}_\theta^e(K)$), the agent decides whether to post the query to the public platform. Posting entails a query-type-specific cost $d \sim \Gamma_\theta$ with CDF $\Gamma_\theta : [0,\infty) \to [0,1]$; this cost captures time to format a question, privacy concerns, stigma, and verification costs (Kummer, Slivko, and Zhang 2020). The agent's expected benefit from posting is

$$U_\theta^e(K) \equiv \sigma^e(K) \cdot V_\theta, \tag{3}$$

where $V_\theta > 0$ is the private value of obtaining a public answer and $\sigma^e(K) \in [0,1]$ is the lifetime resolution probability of a posted query at state K in environment e (derived from the answering game below). The agent posts iff $U_\theta^e(K) \ge d$, implying an endogenous escalation probability

$$m_\theta^e(K) \equiv \Gamma_\theta\big(U_\theta^e(K)\big) \in [0,1], \tag{4}$$

and therefore a posted flow of type-θ queries

$$q_\theta^e(K) \equiv m_\theta^e(K)\big[1 - a_\theta^e(K)\big]. \tag{5}$$

The total posted flow is $Q^e(K) = \pi\, q_L^e(K) + (1-\pi)\, q_H^e(K)$. The object $q_H^e(K)$ is the *matching-to-the-commons margin*: it is the flow of knowledge-enhancing problems that the public platform has a chance to work on, equal to the fraction of type-H queries that fail private resolution times the fraction of those failures that are escalated to the platform.

2.4 Answering on the public platform

A posted query remains active for $T \geq 1$ periods; the baseline analysis sets $T = 1$, so that a query is either resolved within the period it is posted or it expires (Appendix B discusses the general case $T > 1$, where the qualitative results carry through). Each participating agent can process at most one query per period. With $\Psi(\alpha^{*,e}(K))$ participating agents and $Q^e(K)$ posted queries (under $T = 1$, this is also the stock of active queries), the probability that a given active query is matched to an agent is

$$\mu^e(K) = \min\left\{1,\ \frac{\Psi(\alpha^{*,e}(K))}{Q^e(K)}\right\}. \tag{6}$$

When the contributor pool is large relative to posted flow ($\Psi(\alpha^{*,e}) \geq Q^e$), every query is matched and $\mu = 1$; when the pool is small ($\Psi(\alpha^{*,e}) < Q^e$), queries compete for scarce contributor attention and $\mu < 1$. This capacity constraint is what makes pool thinning mechanically lower resolution.[3] If no queries are posted ($Q^e(K) = 0$), we define $\mu^e(K) = 0$, $\sigma^e(K) = 0$, $\bar{\Delta}^e(K) = 0$, and $S^e(K) = 0$.

Conditional on being matched, the assigned agent draws an i.i.d. answering cost $c \sim F$ on $[0, \bar{c}]$, independent of the agent's ability and of the query's type and residual difficulty. The cost c measures idiosyncratic match quality between agent and query: how well the agent's specific knowledge, tooling, and context fit the posted question, which is conceptually distinct from residual difficulty r that measures amenability to private resolution.[4] For example, a question about a deprecated COBOL library may have high residual difficulty r (no modern AI or documentation covers it well) but draw low c if matched to a retired developer who maintained that library, and high c if matched to a leading computer scientist who has never touched COBOL. Conversely, a technically deep question about concurrent data structures may draw low c for a systems programmer who works on exactly that problem, despite the question being objectively hard.

3. The $\min\{1, \Psi/Q\}$ form is a short-side (Leontief) matching rule that assumes all capacity is perfectly allocated to queries without search frictions. Replacing it with $M(\Psi, Q)/Q$ for a standard aggregate matching function M would add search frictions, ensuring some queries go unmatched even when $\Psi > Q$. The qualitative comparative statics are unchanged: pool thinning lowers matching rates under any matching technology with $M_\Psi > 0$. See Appendix C for further discussion.

4. The assumption that c is drawn from the same distribution F regardless of query type is a substantive modeling choice; Appendix C.2 discusses the interpretation and extensions with type-indexed costs.

The matched agent observes the posted query but does not condition their answering decision on its type classification. Under $T = 1$, the query expires at the end of the period, so the agent faces a one-shot answering decision with no option value of waiting. If the agent answers, they incur cost $C(K) + c$ (where $C(K)$ is a state-dependent cost shifter with $C'(K) < 0$, capturing that a richer archive makes answering cheaper) and receive $\beta \bar{\Delta}^e(K) + u$, where $u \geq 0$ is a private benefit from answering (intrinsic motivation, reciprocity, reputation) and $\beta \geq 0$ is the agent's concern for the archive expansion their answer generates. The expected knowledge increment from answering a randomly drawn posted query is

$$\bar{\Delta}^e(K) = \frac{(1-\pi)\, q_H^e(K) \cdot \Delta}{Q^e(K)}, \tag{7}$$

which equals $\Delta > 0$ (the knowledge increment from a resolved type-H query) times the probability that the posted query is type H. The parameter β captures general concern for the growth of the public archive: professional norms, community identity, internalized social value, or direct consumption value from a richer commons. Because the archive is a non-excludable public good in a continuum of agents, β should be interpreted as an internalized or perceived value of contributing, not as a rational calculation of marginal impact on the aggregate state.

The agent answers iff the net benefit is non-negative, yielding an answering cutoff in closed form:

$$c^{*,e}(K) = \max\{0,\ \beta \bar{\Delta}^e(K) + u - C(K)\}. \tag{8}$$

Under $T = 1$, the lifetime resolution probability of a posted query coincides with the per-period resolution hazard:

$$\sigma^e(K) = \mu^e(K) \cdot F\big(c^{*,e}(K)\big). \tag{9}$$

Two factors determine σ: whether the query is matched to a contributor (extensive margin, μ^e) and whether the matched contributor answers (intensive margin, $F(c^{*,e})$). AI can reduce $\sigma^e(K)$ through pool thinning: when $\alpha^{*,e}(K)$ falls, $\Psi(\alpha^{*,e}(K))$ falls, which lowers $\mu^e(K)$ if the platform is capacity-constrained.

2.5 Participation surplus

A participating agent is matched to a posted query with probability $\min\{1, Q^e(K)/\Psi(\alpha^{*,e}(K))\}$ per period: when $Q \geq \Psi$, every participant is matched; when $Q < \Psi$, queries are scarce and not all participants receive one. Conditional on being matched, the agent's expected surplus from the answering opportunity is

$$\Pi^e(K) = \int_0^{c^{*,e}(K)} \left(c^{*,e}(K) - c\right) dF(c), \tag{10}$$

which equals zero when $c^{*,e}(K) = 0$ (no agent answers). The expected per-period payoff from participation is

$$S^e(K) = \min\left\{1, \frac{Q^e(K)}{\Psi(\alpha^{*,e}(K))}\right\} \cdot \Pi^e(K). \tag{11}$$

The ratio Q^e/Ψ in (11) is inverted relative to the match probability $\mu^e = \min\{1, \Psi/Q^e\}$: μ measures the probability a posted query encounters a contributor, while $\min\{1, Q^e(K)/\Psi(\alpha^{*,e}(K))\}$ is the probability a participating agent is matched to a query. The surplus S^e does not depend on α: conditional on participating, all agents face the same matching and answering game. The non-excludable archive benefit (K improves everyone's tools) accrues to all agents regardless of participation and therefore does not enter the participation comparison. Heterogeneity in the participation decision is driven entirely by heterogeneity in outside options $w^e(\alpha, K; \pi)$.

2.6 Knowledge accumulation and timing

Public knowledge accumulates only from knowledge-enhancing (type-H) problems whose solutions are publicly logged. In the human-only benchmark, this occurs only when a posted type-H query is resolved on the platform. In the AI regime, public knowledge can additionally grow when a privately produced AI solution is converted into a public artifact at rate $\eta \in [0, 1]$: when private resolution succeeds for a type-H query, the solution enters the public archive with probability η. Expected new public knowledge at state K is

$$h^{\mathrm{HO}}(K) = \Delta(1 - \pi)\, q_H^{\mathrm{HO}}(K)\, \sigma^{\mathrm{HO}}(K), \tag{12}$$

and

$$h^{\mathrm{AI}}(K;\eta) = \Delta(1-\pi)\Big[q_H^{\mathrm{AI}}(K)\,\sigma^{\mathrm{AI}}(K) + \eta\, a_H^{\mathrm{AI}}(K)\Big]. \tag{13}$$

The first term is human-generated public knowledge from posted queries resolved on the platform; the second term is privately resolved AI knowledge converted to the commons. The public archive stock evolves according to

$$K_{t+1} = (1-\lambda)K_t + h^e(K_t;\eta). \tag{14}$$

A steady state K^e satisfies $h^e(K^e;\eta) = \lambda K^e$. Define average creation $\phi^{\mathrm{AI}}(K) \equiv h^{\mathrm{AI}}(K;0)/K$ for $K > 0$. The decomposition that anchors the paper is immediate from (13): in the AI regime, created public knowledge equals posted flow of knowledge-enhancing queries times resolution conditional on posting, plus a conversion term for private AI successes. The benchmark case $\eta = 0$, in which private AI resolutions do not enter the public archive, isolates the crowd-out mechanism; Section 5 studies how increasing η shifts creation and changes long-run outcomes.

The within-period timing is as follows. In Stage 0 (participation), each agent observes K_t and the environment e and participates iff $S^e(K_t) \geq w^e(\alpha_i, K_t; \pi)$, determining the contributor pool $\Psi(\alpha^{*,e}(K_t))$. In Stage 1 (query arrival, private resolution, posting), each agent draws a task, attempts private resolution, and decides whether to post, determining posted flows q_L^e, q_H^e, and total flow Q^e. In Stage 2 (matching and answering), active queries are matched to participating agents under the congestion technology, matched agents draw $c \sim F$ and answer iff $c \leq c^{*,e}(K_t)$, and resolved queries are publicly logged. In Stage 3 (accumulation), resolved type-H queries contribute Δ to K; under AI with $\eta > 0$, each privately resolved type-H query contributes Δ with probability η; and the archive updates to K_{t+1}.

2.7 Period equilibrium

Given (K, e), a period equilibrium is a tuple $(\alpha^*, c^*, \sigma, \mu, q_L, q_H, Q, S, \bar{\Delta})$ satisfying the following conditions simultaneously:

(E1) *Matching probability.* $\mu^e(K) = \min\{1, \Psi(\alpha^{*,e}(K))/Q^e(K)\}$, with $\mu^e = 0$ if $Q^e = 0$.

(E2) *Posted-pool composition.* $\bar{\Delta}^e(K) = (1-\pi)\, q_H^e(K) \cdot \Delta / Q^e(K)$ if $Q^e > 0$; $\bar{\Delta}^e = 0$ if $Q^e = 0$.

(E3) *Answering cutoff.* $c^{*,e}(K) = \max\{0,\ \beta\,\bar{\Delta}^e(K) + u - C(K)\}$.

(E4) *Lifetime resolution probability.* $\sigma^e(K) = \mu^e(K) \cdot F(c^{*,e}(K))$.

(E5) *Posting benefit and escalation.* $U^e_\theta(K) = \sigma^e(K) \cdot V_\theta$ and $m^e_\theta(K) = \Gamma_\theta(U^e_\theta(K))$ for $\theta \in \{L, H\}$.

(E6) *Posted flows.* $q^e_\theta(K) = m^e_\theta(K) \cdot [1 - a^e_\theta(K)]$ and $Q^e(K) = \pi\, q^e_L(K) + (1-\pi)\, q^e_H(K)$.

(E7) *Participation surplus.* $S^e(K) = \min\{1, Q^e(K)/\Psi(\alpha^{*,e}(K))\} \cdot \Pi^e(K)$.

(E8) *Participation cutoff.* $S^e(K) = w^e(\alpha^{*,e}(K), K; \pi)$.

The answering cutoff $c^{*,e}$ is in closed form given $\bar{\Delta}^e$ and K. The remaining fixed point has a nested structure. The *inner loop*, given α^* (and hence the contributor pool $\Psi(\alpha^*)$), solves for posted flows (q_L, q_H): these depend on σ^e, which depends on $\mu^e = \min\{1, \Psi(\alpha^*)/Q^e\}$ and on $c^* = \max\{0, \beta\bar{\Delta}^e + u - C(K)\}$, both of which depend on Q^e through $\bar{\Delta}^e$ and the congestion ratio. This is a fixed point in (q_L, q_H) at given α^*. The *outer loop* finds α^* such that $S^e(K) = w^e(\alpha^*, K; \pi)$, where S^e is computed from the inner-loop solution. A steady state additionally requires $h^e(K^e; \eta) = \lambda K^e$, which pins down the long-run archive stock.

The inner loop can admit multiple fixed points in (q_L, q_H) for a given α^*, through self-reinforcing expectations about posted-pool composition: if agents expect the posted pool to be H-rich, the answering cutoff c^* is high, which raises σ, which encourages posting, and if the benefit from public answers is higher for H-type queries ($V_H > V_L$ with Γ_H sufficiently elastic), the additional posting is disproportionately H-type, confirming the high $\bar{\Delta}^e$. This within-period coordination problem is conceptually distinct from the across-period accumulation dynamics in K that drive the main results; where relevant, we select the Pareto-dominant inner-loop fixed point. Appendix C provides formal conditions under which the inner loop has a unique solution. For the dynamic results in Sections 3–5, we maintain that the selected branch of inner-loop equilibria varies continuously with K. Under the uniqueness conditions in Appendix C, this continuity follows from the primitives by the implicit function theorem; when the inner loop is multiple, it is a maintained property of the selection rule.

3 AI and the Two Margins

This section derives the paper's main comparative statics under the benchmark case $\eta = 0$, in which private AI resolutions do not enter the public archive. We focus on type-H queries, since

only H contributes to the archive stock. The results are organized around the decomposition in (13): AI affects archive creation $h^{\mathrm{AI}}(K;0) = \Delta(1-\pi)\, q_H^{\mathrm{AI}}(K)\, \sigma^{\mathrm{AI}}(K)$ through two equilibrium objects: the posted flow of knowledge-enhancing queries that reach the platform, $q_H^e(K)$ (the *flow margin*), and the resolution probability conditional on posting, $\sigma^e(K)$ (the *resolution margin*). Because both are equilibrium objects, they are not causally independent: $q_H^e = m_H^e(1-a_H^e)$ depends on the resolution probability σ^e through the endogenous posting decision m_H^e, and σ^e depends on posted flow through the congestion ratio Ψ/Q^e. The decomposition is therefore an accounting separation of archive creation into posted flow and conditional resolution, not a one-for-one mapping between causal mechanisms and observable objects. The primary mechanism behind the flow margin is private diversion (AI resolves queries off-platform); the primary mechanism behind the resolution margin is pool thinning (contributor exit reduces matching capacity). We begin with a composition result that isolates one channel through which AI can improve the intensive margin of resolution, and then show that a competing extensive-margin channel (pool thinning through congestion) can overwhelm it.

3.1 Composition of the posted pool

The answering cutoff $c^{*,e}(K) = \max\{0,\, \beta\bar{\Delta}^e(K) + u - C(K)\}$ from (8) depends on the environment e only through the expected knowledge increment $\bar{\Delta}^e(K)$, which in turn depends on the H-share of the posted pool. If AI changes what is posted, it changes the incentive to answer, even holding the contributor pool fixed. The following result isolates this composition channel.

Lemma 1 (Composition of the posted pool and the answering cutoff)**.** *Fix K. The expected knowledge increment satisfies*

$$\bar{\Delta}^e(K) = \Delta \cdot \frac{(1-\pi)\, q_H^e(K)}{\pi\, q_L^e(K) + (1-\pi)\, q_H^e(K)}, \tag{15}$$

which is increasing in the ratio q_H^e/q_L^e. Therefore $c^{*,\mathrm{AI}}(K) \geq c^{*,\mathrm{HO}}(K)$ iff the posted pool is weakly more H-rich under AI:

$$\frac{q_H^{\mathrm{AI}}(K)}{q_L^{\mathrm{AI}}(K)} \geq \frac{q_H^{\mathrm{HO}}(K)}{q_L^{\mathrm{HO}}(K)}. \tag{16}$$

A sufficient condition, holding escalation probabilities fixed ($m_\theta^{\mathrm{AI}} = m_\theta^{\mathrm{HO}}$ for both θ), is that AI

disproportionately resolves routine queries privately:

$$\frac{1-a_H^{\mathrm{AI}}(K)}{1-a_H^{\mathrm{HO}}(K)} > \frac{1-a_L^{\mathrm{AI}}(K)}{1-a_L^{\mathrm{HO}}(K)}, \tag{17}$$

i.e., the private-failure rate for H-type queries declines proportionally less than for L-type queries. This holds when AI is relatively more capable at routine tasks than at knowledge-enhancing ones.

The expression for $\bar{\Delta}^e$ is Δ times the H-share of the posted pool. Under fixed escalation, $q_\theta^e = m_\theta \cdot (1-a_\theta^e)$, so $q_H^e/q_L^e = (m_H/m_L)\cdot(1-a_H^e)/(1-a_L^e)$; with m_H/m_L held fixed, this ratio rises under AI iff (17) holds. The sufficient condition is a partial-equilibrium statement: in equilibrium, m_θ^e depends on σ^e, which depends on the participation cutoff and on c^* itself. If AI's differential capability across types is strong enough, however, it enriches the posted pool even after accounting for equilibrium feedback.

Lemma 1 connects to the selection logic emphasized by J. S. Gans (2024): when AI diverts easy (routine) queries to private resolution, the posted pool shifts toward harder, knowledge-enhancing problems, raising the expected public-good benefit from answering and thereby increasing willingness to answer. The composition channel here operates through the answering cutoff c^* (intensive margin), isolated from a separate congestion channel that operates through the match probability μ (extensive margin); the interaction between the two is the subject of Proposition 1. The composition channel requires $\beta > 0$: when agents do not value archive expansion ($\beta = 0$), the answering cutoff $c^* = \max\{0, u - C(K)\}$ is independent of the posted-pool composition, and force (i) in Proposition 1 is absent.

3.2 Resolution: composition vs. congestion

Lemma 1 shows that AI can raise the answering cutoff through posted-pool composition. Whether this translates into higher resolution probability depends on the extensive margin: the match probability $\mu^e(K)$ that determines whether a posted query encounters a contributor at all. In the congested regime ($\Psi(\alpha^{*,e}) < Q^e$), both the composition and the congestion channels are active, and the sign of $\sigma^{\mathrm{AI}} - \sigma^{\mathrm{HO}}$ is the outcome of a race.

Proposition 1 (Resolution: composition vs. congestion)**.** *Fix K and suppose the platform is congested in both environments: $\Psi(\alpha^{*,e}(K)) < Q^e(K)$ for $e \in \{\mathrm{HO}, \mathrm{AI}\}$. Under $T = 1$, the lifetime*

resolution probability is

$$\sigma^e(K) = \frac{\Psi(\alpha^{*,e}(K))}{Q^e(K)} \cdot F\big(c^{*,e}(K)\big). \tag{18}$$

Resolution falls under AI ($\sigma^{\mathrm{AI}}(K) < \sigma^{\mathrm{HO}}(K)$) whenever pool thinning dominates the combined effect of the composition shift and posted-flow reduction:

$$\frac{\Psi(\alpha^{*,\mathrm{AI}}(K))}{\Psi(\alpha^{*,\mathrm{HO}}(K))} \cdot \frac{F\big(c^{*,\mathrm{AI}}(K)\big)}{F\big(c^{*,\mathrm{HO}}(K)\big)} < \frac{Q^{\mathrm{AI}}(K)}{Q^{\mathrm{HO}}(K)}. \tag{19}$$

Three forces shape the resolution comparison, two of which raise σ and one of which lowers it. First, the *composition* channel (Lemma 1): AI selectively diverts routine queries, enriching the posted pool, which raises $\bar{\Delta}^e$ and hence c^* and $F(c^*)$. Second, *posted-flow reduction*: AI lowers Q^e by resolving more queries privately, and in the congested regime fewer posted queries competing for the same pool raises the match probability $\mu^e = \Psi/Q$. Third, *pool thinning*: AI raises outside options ($w^{\mathrm{AI}} > w^{\mathrm{HO}}$), which lowers $\alpha^{*,e}$ and reduces the contributor pool $\Psi(\alpha^{*,e})$, which in the congested regime lowers μ^e. The condition in (19) simply factors the ratio $\sigma^{\mathrm{AI}}/\sigma^{\mathrm{HO}}$ into these three components: the pool ratio (less than one), the $F(c^*)$ ratio (weakly greater than one under Lemma 1), and the inverse posted-flow term $Q^{\mathrm{HO}}/Q^{\mathrm{AI}}$ (greater than one, since lower posted flow under AI eases congestion and raises the match probability). Resolution falls when the pool decline is large enough relative to the other two adjustments. The qualitative race among these three forces extends to any matching technology with $M_\Psi > 0$ and to general $T \geq 1$: pool thinning lowers μ, composition enrichment raises $F(c^*)$, and the sign of $\sigma^{\mathrm{AI}} - \sigma^{\mathrm{HO}}$ remains determined by which force dominates. The specific factored form in (19) uses the Leontief structure.

Remark 1 (Discriminating prediction: supply vs. demand shocks). The factored form in (18) implies a discriminating prediction for the source of resolution decline. In the congested regime, a pure demand reduction (Q^e falls while $\Psi(\alpha^{*,e})$ is unchanged) raises the match probability Ψ/Q and therefore pushes σ upward: fewer queries competing for the same contributor pool means each posted query is more likely to encounter an answerer. A pure supply reduction ($\Psi(\alpha^{*,e})$ falls while Q^e is unchanged) lowers Ψ/Q and pushes σ downward. This asymmetry holds under any matching technology with $M_\Psi > 0$ and under general $T \geq 1$, since $\sigma = 1 - (1 - \mu \cdot F(c^*))^T$ is increasing in μ for any T. The asymmetry means that observing resolution decline alongside posted-flow decline requires that pool thinning is quantitatively present and dominant: improved

private resolution alone (a demand-side shock that diverts questions from the platform) would ease congestion and raise σ, not lower it. This is consistent with the micro-level evidence of Li and Kim (2025) that high-activity answerers on Stack Overflow disproportionately reduce contributions after ChatGPT's release, supporting the supply-side pool-thinning channel in Proposition 1. Section 6 discusses quality-composition mechanisms that go beyond the baseline model's headcount and congestion channel.

The empirical counterpart is the "difficulty up, volume down, resolution down" pattern: AI diverts routine questions from the posted pool (Lemma 1), posted volume falls, remaining questions become harder, yet resolution declines because the contributor pool has thinned (force (iii) dominates). This pattern is consistent with the Stack Overflow evidence documented by Del Rio-Chanona, Laurentsyeva, and Wachs (2024); Remark 1 notes that the simultaneous decline in both volume and resolution is diagnostic of the supply-side channel.

The comparison with J. S. Gans (2024) clarifies what the present architecture adds. J. S. Gans (2024) identifies the core economic intuition: AI strengthens the private outside option, truncating the posted pool and changing incentive structures on the platform. In his static model, truncation yields a clean sign result: the remaining posted queries are harder, so contributors are more willing to answer. Proposition 1 shows that when contributors are heterogeneous and participation is endogenous, the same truncation that enriches the posted pool (Lemma 1) coexists with pool thinning that reduces matching capacity, and the sign of $\sigma^{\mathrm{AI}} - \sigma^{\mathrm{HO}}$ becomes the outcome of a race rather than a clean comparative static. The posted-flow channel is the same force he identified; what the present model adds is the contributor-side response (pool thinning through improved outside options, which his static setting does not feature) and, in the next section, the dynamic interaction between archive quality and participation that generates multiple steady states.

3.3 Dynamic crowd-out

We now use the two margins jointly to characterize long-run archive outcomes. At the human-only steady state K^{HO}, creation exactly offsets depreciation: $h^{\mathrm{HO}}(K^{\mathrm{HO}}) = \lambda K^{\mathrm{HO}}$. If AI reduces creation at that point, the archive cannot be sustained.

Proposition 2 (Dynamic crowd-out and lower archive steady states)**.** *Let $K^{\mathrm{HO}} > 0$ be a steady*

state in the human-only economy: $h^{\mathrm{HO}}(K^{\mathrm{HO}}) = \lambda K^{\mathrm{HO}}$. *Suppose at* K^{HO} *the AI economy satisfies*

$$h^{\mathrm{AI}}(K^{\mathrm{HO}}; 0) < h^{\mathrm{HO}}(K^{\mathrm{HO}}), \tag{20}$$

i.e., AI reduces one-step knowledge creation at the human-only steady state (under $\eta = 0$*). Then* $h^{\mathrm{AI}}(K^{\mathrm{HO}}; 0) - \lambda K^{\mathrm{HO}} < 0$*. If additionally* $h^{\mathrm{AI}}(0; 0) \geq 0$ *and* $h^{\mathrm{AI}}(\cdot\,; 0)$ *is continuous, the AI economy admits a steady state* $K^{\mathrm{AI}} \in [0, K^{\mathrm{HO}})$*.*

Proposition 2 is an existence result: when AI reduces knowledge creation at the human-only benchmark, the intermediate value theorem delivers a lower AI steady state. The proof is in Appendix A. The continuity of $h^{\mathrm{AI}}(\cdot\,; 0)$ follows from the maintained assumption in Section 2.7 that the selected branch of inner-loop equilibria varies continuously with K. [5] Note Proposition 2 does not assume that resolution conditional on posting improves under AI; both margins can decline simultaneously, which matches the empirical pattern more closely. The same IVT logic extends to general T, conditional on continuity of the selected branch and the benchmark inequality $h^{\mathrm{AI}}(K^{\mathrm{HO}}; 0) < h^{\mathrm{HO}}(K^{\mathrm{HO}})$.

The following corollary decomposes the condition in Proposition 2 into the two margins and connects them to the channels identified in Proposition 1.

Corollary 1 (Two-margin decomposition of crowd-out)**.** *The condition* $h^{\mathrm{AI}}(K^{\mathrm{HO}}; 0) < h^{\mathrm{HO}}(K^{\mathrm{HO}})$ *is equivalent to*

$$q_H^{\mathrm{HO}}(K^{\mathrm{HO}})\, \sigma^{\mathrm{HO}}(K^{\mathrm{HO}}) > q_H^{\mathrm{AI}}(K^{\mathrm{HO}})\, \sigma^{\mathrm{AI}}(K^{\mathrm{HO}}). \tag{21}$$

The decline decomposes as

$$q_H^{\mathrm{HO}} \sigma^{\mathrm{HO}} - q_H^{\mathrm{AI}} \sigma^{\mathrm{AI}} = \underbrace{\bar{\sigma}\, \Delta q_H}_{\text{flow margin}} + \underbrace{\bar{q}_H\, \Delta \sigma}_{\text{resolution margin}}, \tag{22}$$

where $\Delta q_H \equiv q_H^{\mathrm{HO}} - q_H^{\mathrm{AI}}$*,* $\Delta \sigma \equiv \sigma^{\mathrm{HO}} - \sigma^{\mathrm{AI}}$*, and bars denote averages across regimes.*[6] *The flow margin captures the change in posted H-type volume; the resolution margin captures the change in*

5. If the inner loop admits multiple fixed points for some values of K, the Pareto-dominant selection may introduce discontinuities in h^{AI} at points where the set of inner-loop equilibria changes. The maintained assumption in Section 2.7 rules out such transitions within the relevant range $[0, K^{\mathrm{HO}}]$.

6. The identity follows from $\bar{\sigma}\, \Delta q_H + \bar{q}_H\, \Delta \sigma = \frac{1}{2}(q_H^{\mathrm{HO}} \sigma^{\mathrm{HO}} - q_H^{\mathrm{AI}} \sigma^{\mathrm{HO}} + q_H^{\mathrm{HO}} \sigma^{\mathrm{AI}} - q_H^{\mathrm{AI}} \sigma^{\mathrm{AI}}) + \frac{1}{2}(q_H^{\mathrm{HO}} \sigma^{\mathrm{HO}} - q_H^{\mathrm{HO}} \sigma^{\mathrm{AI}} + q_H^{\mathrm{AI}} \sigma^{\mathrm{HO}} - q_H^{\mathrm{AI}} \sigma^{\mathrm{AI}}) = q_H^{\mathrm{HO}} \sigma^{\mathrm{HO}} - q_H^{\mathrm{AI}} \sigma^{\mathrm{AI}}$. Midpoint weighting absorbs the interaction term that appears under base-period weighting.

the probability that posted queries are resolved. The condition in Proposition 2 holds whenever the sum of the two margins is positive.

The decomposition is an accounting identity, not a causal decomposition: it describes *how* archive creation declines but does not separately identify why each margin moved. A decline in the flow component ($\Delta q_H > 0$) is most naturally associated with stronger private diversion, but because posting is endogenous to σ^e, it can also reflect equilibrium feedback through resolution and congestion. The empirical counterpart is the decline in question volume documented by Del Rio-Chanona, Laurentsyeva, and Wachs (2024) and the traffic diversion documented by Padilla et al. (2025). A decline in the resolution component ($\Delta\sigma > 0$) reflects the net outcome of the race in Proposition 1: pool thinning, composition enrichment, and congestion relief jointly determine whether posted queries are more or less likely to be resolved. Importantly, the two components respond to different instruments: public logging ($\eta > 0$, studied in Section 5) offsets the decline associated with private diversion by converting off-platform resolutions into public artifacts, but it does not repair congestion- or participation-driven deterioration in conditional resolution, which requires maintaining contributor engagement on the platform.

4 Dynamic Crowd-Out and the Self-Undermining Feedback

Proposition 2 established that AI can lower the long-run archive stock when it reduces knowledge creation at the human-only steady state. This section studies a further amplification mechanism: the interaction between archive quality and contributor participation through outside options. The mechanism rests on a property of the outside-option function that was maintained but not yet exploited: $w_K^e > 0$, i.e., a richer archive raises outside options by improving retrieval-augmented and AI-assisted tools that complement agents' own work. This creates a self-undermining feedback loop: as the archive grows, outside options improve, contributors exit the platform, and the resulting pool thinning slows archive growth. When this feedback is strong enough, the average creation rate is hump-shaped in K, generating multiple steady states and a structural minimum viable archive below which the platform cannot sustain itself. The structural threshold is not AI-specific: both the pre-AI and AI economies possess it, because the empty-archive shutdown (Assumption 1(a) below) is environment-independent. What AI does is compress the stable steady state inward, narrowing the viable region through the same two margins

identified in the previous section.

4.1 The self-undermining feedback

The feedback operates through the participation margin. At the participation cutoff, $S^e(K) = w^e(\alpha^{*,e}(K), K; \pi)$, where $S^e(K)$ is the expected per-period payoff from volunteering and w^e is the outside option. When K increases, two competing effects arise on the contributor side. The first is an answering-cost reduction: $C'(K) < 0$ means a richer archive makes answering cheaper, which raises the answering cutoff $c^{*,e}(K)$, which raises both $\sigma^e(K)$ (making posting more attractive) and $\Pi^e(K)$ (making volunteering more rewarding). The second is outside-option improvement: $w^e_K > 0$ means a richer archive also raises the return to own work, pulling agents away from the platform. The first effect tends to sustain platform activity; the second tends to undermine it. Which effect dominates depends on K.

A partial stabilizer is available in the uncongested regime ($\Psi(\alpha^{*,e}) > Q^e$), where not all participants are matched: as Ψ falls toward Q, each remaining participant is matched more often ($\min\{1, Q/\Psi\}$ rises), which raises $S^e(K)$ and partially offsets exit pressure. Once the platform tips into congestion ($\Psi < Q$), however, this stabilizer is exhausted: every participant is already matched with probability one, so further thinning of the pool has no offsetting effect on participant match rates. The asymmetry creates a natural non-linearity: the platform is resilient to moderate exit while uncongested but becomes fragile once congested.

4.2 Multiple steady states

The following result shows that the self-undermining feedback, combined with low-K scarcity, can generate multiple steady states in the AI economy. Define average creation $\phi^{\mathrm{AI}}(K) \equiv h^{\mathrm{AI}}(K; 0)/K$ for $K > 0$. Rather than assuming a shape for ϕ^{AI}, we state conditions on the model's primitives that deliver the required shape through the participation and congestion mechanisms.

Assumption 1 (Primitive conditions for multiple steady states)**.** Maintain $\eta = 0$.

(a) **Empty-archive shutdown:** $C(0) > \beta\Delta + u$. When the archive is empty, the answering cost shifter is so high that no agent is willing to answer even the most valuable query: $\beta\bar{\Delta}^e(0) + u - C(0) \leq \beta\Delta + u - C(0) < 0$, so $c^{*,e}(0) = 0$.

(b) **Platform viability at intermediate K:** There exists $\hat{K} > 0$ at which the model's primitives jointly support a functioning platform with $\phi^{\text{AI}}(\hat{K}) > \lambda$. That is, at $\hat{K}$ the answering cost $C(\hat{K})$ is low enough that $c^* > 0$, the contributor pool $\Psi(\alpha^*)$ and posted flow Q are both positive, the resulting resolution probability σ is high enough to sustain posting, and the product $h^{\text{AI}}(\hat{K}; 0)$ exceeds $\lambda\hat{K}$. Unlike (a) and (c), this condition cannot be expressed as a single inequality on primitive parameters; it requires that the functional forms jointly deliver a viable platform at some $\hat{K}$, which is verified in the parametric example (Appendix C.5).

(c) **Unbounded outside options:** $\lim_{K\to\infty} w^{\text{AI}}(\underline{\alpha}, K; \pi) = \infty$. Even the lowest-ability agent eventually exits the platform as K grows, because the return to own work grows without bound.

Additionally, under the maintained selection assumption in Section 2.7, ϕ^{AI} is continuous on $(0, \infty)$.

Condition (a) implies low-K scarcity: when $C(0) > \beta\Delta + u$, continuity of C ensures $c^{*,e}(K) = 0$ in a neighborhood of $K = 0$, so $F(c^*) = 0$, $\sigma^e = 0$, the posting benefit $U_\theta^e = 0$, and therefore $m_\theta^e = \Gamma_\theta(0) = 0$ (since Γ_θ admits a density). All activity collapses: $Q^e = 0$, $h^{\text{AI}} = 0$, and $\phi^{\text{AI}}(K) \to 0 < \lambda$ as $K \downarrow 0$. Condition (c) implies high-K drain: when $w^{\text{AI}}(\underline{\alpha}, K; \pi) \to \infty$, the participation surplus $S^e(K)$ (which is bounded by $\Pi^e(K) \le \bar{c}$) eventually falls below $w^{\text{AI}}(\underline{\alpha}, K; \pi)$ for all $\underline{\alpha}$, so $\alpha^{*,e}(K) < \underline{\alpha}$ and $\Psi(\alpha^{*,e}) = 0$ for all sufficiently large K. With no contributors, $\mu^e = 0$, $\sigma^e = 0$, $h^{\text{AI}} = 0$, and $\phi^{\text{AI}}(K) \to 0 < \lambda$. Condition (b) provides the existence requirement. Together, the three conditions imply that ϕ^{AI} is below λ near zero, above λ at some interior $\hat{K}$, and below λ for large K, which is the shape that generates multiple steady states.

The distinction from the aggregate-effort framework of Acemoglu, Kong, and Ozdaglar (2026) is worth noting. In their model, knowledge decline follows from a shape assumption on the relationship between AI accuracy and human effort. Here, the hump in ϕ^{AI} arises from the interaction between archive quality and contributor exit through the congestion matching technology: (a) reflects that answering requires archive support, (c) reflects that archive quality eventually pulls contributors away, and (b) reflects that productive matching is possible in between.

Proposition 3 (Self-undermining feedback and multiple steady states)**.** *Under Assumption 1, $\phi^{\text{AI}}(\cdot)$ crosses λ at least twice, and the AI economy has at least two positive steady states $0 < K_U^{\text{AI}} <$*

K_H^{AI}, where K_U^{AI} is an unstable threshold and K_H^{AI} is locally stable. If $K_t < K_U^{\mathrm{AI}}$, the archive declines toward zero or a lower steady state.

The proof, provided in Appendix A, follows from the intermediate value theorem applied to $h^{\mathrm{AI}}(K;0) - \lambda K = K(\phi^{\mathrm{AI}}(K) - \lambda)$: the function is negative near zero (from (a)), positive at $\hat{K}$ (from (b)), and negative for large K (from (c)), so it crosses zero at least twice. Sign changes determine stability: K_U^{AI} is unstable (net growth is negative just below and positive just above) and K_H^{AI} is locally stable (net growth is positive just below and negative just above). The threshold K_U^{AI} acts as a basin boundary: platforms whose archive stock falls below K_U^{AI} experience self-reinforcing decline because creation falls short of depreciation, while platforms above K_U^{AI} converge to the high steady state. The primitive conditions in Assumption 1 use the $T = 1$ structure for the specific expressions; under $T > 1$ the same qualitative conditions hold but the threshold expressions change (Appendix B). The two-crossing structure is not AI-specific: the empty-archive shutdown in condition (a) depends on $C(0) > \beta\Delta + u$, which is environment-independent, and the AI-specific terms ($\gamma_w K$, ρK) vanish as $K \to 0$, so both economies activate at approximately the same archive stock. The human-only economy satisfies the same conditions whenever w^{HO} grows without bound in K (condition (c) with $\gamma_w^{\mathrm{HO}} > 0$); even when $\gamma_w^{\mathrm{HO}} = 0$, as in the parametric example, ϕ^{HO} can exhibit a low-K threshold of similar magnitude. The AI-specific effect is the compression of the stable steady state: $K_H^{\mathrm{AI}} < K_H^{\mathrm{HO}}$ (Proposition 2), which narrows the viable region.

The economic logic is as follows. At very low K, the archive is too thin to support productive matching: answering costs are high, resolution is low, and the platform cannot generate enough new knowledge to offset depreciation. As K rises into the intermediate range, answering becomes cheaper, resolution improves, and the platform enters a region of net growth. But as K continues to grow, the self-undermining feedback intensifies: outside options improve ($w_K^e > 0$), the participation cutoff falls, the contributor pool thins, and eventually the congestion effect overwhelms the answering-cost reduction. The platform's own success in building the archive raises the opportunity cost of contributing to it.[7]

7. A natural question is: if the archive declines toward a low-K region, why not reverse the process by reducing reliance on AI? In practice this is not a frictionless margin. AI is an off-platform outside option that individual users and contributors cannot unilaterally remove, and platform-wide reduction in AI use would require coordination or external regulation. Moreover, when K is low, answering costs $C(K)$ are high, resolution is low, and few agents find it worthwhile to post or participate; this state dependence means that even if some agents attempted to rely more on the platform, the thin contributor pool and poor resolution would discourage them, slowing recovery.

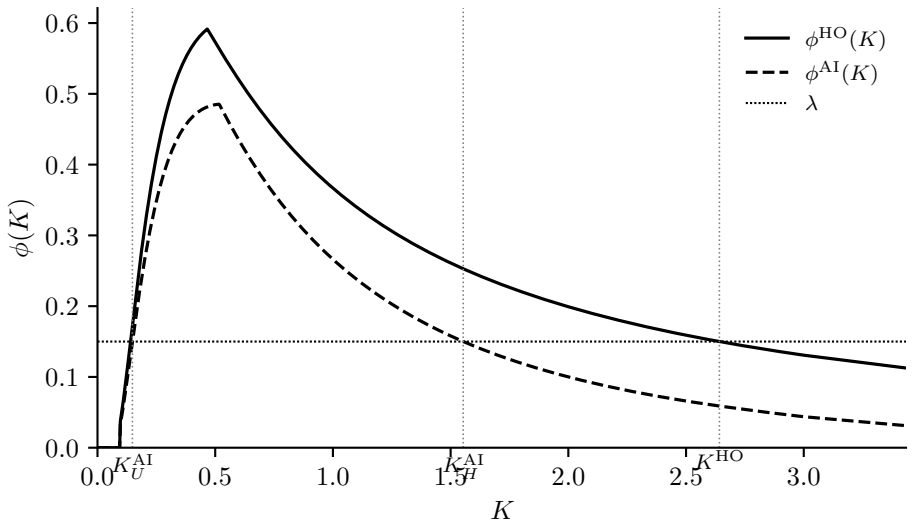


Figure 1: Average knowledge creation rate under the human-only economy (ϕ^{HO}, solid) and the AI economy (ϕ^{AI}, dashed) for the parametric example specified in Appendix C.5, with depreciation rate $\lambda = 0.15$ (densely dotted). Both economies exhibit low-K thresholds at which ϕ crosses λ from below: $K_U^{\mathrm{HO}} \approx 0.14$ and $K_U^{\mathrm{AI}} \approx 0.15$, confirming that the structural minimum viable archive is environment-independent (the empty-archive shutdown depends on $C(0) > \beta\Delta + u$, which does not differ across environments). The AI-specific effect is compression of the stable steady state: $K_H^{\mathrm{AI}} \approx 1.55$ versus $K^{\mathrm{HO}} \approx 2.64$, so the viable region (the interval between the lower crossing and the upper crossing) narrows from approximately 2.50 to 1.41. The hump is lower and narrower under AI because stronger outside options and higher private resolution rates thin the contributor pool and divert queries from the platform.

4.3 Cross-platform vulnerability: the role of π

The routine-task share π enters both margins through which AI reduces archive creation, generating a cross-platform prediction about which types of platforms are most exposed. On the flow side, high-π domains lose more total posted traffic because a larger share of their queries is routine and privately resolved under AI. AI also diverts some H-type queries privately ($a_H^{\mathrm{AI}} > a_H^{\mathrm{HO}}$), reducing the public inflow of knowledge-enhancing problems, although the net response of posted H-flow is an equilibrium outcome rather than a one-for-one consequence of routine-query diversion alone. On the resolution margin, high-π domains experience a

Endogenizing regime choice is a natural extension, but the baseline isolates the externality in the open-archive channel.

larger outside-option shift ($w^{AI} - w^{HO}$ is increasing in π by assumption (iv) in Section 2.2), so the participation cutoff falls more, thinning the pool further. The combined effect makes high-π platforms more exposed overall: they lose more traffic to private resolution and, because the outside-option shift is larger, they lose more contributors to improved outside options. This prediction relies on π governing both query composition and the outside-option shift simultaneously, a feature that distinguishes the present framework from the aggregate-effort model of Acemoglu, Kong, and Ozdaglar (2026) and the static selection model of J. S. Gans (2024), which do not parameterize cross-platform heterogeneity. Platforms like Stack Overflow (high π: much of programming Q&A is routine) are predicted to be more exposed than platforms like MathOverflow (low π: research-level mathematics has a smaller routine share), which is consistent with the differential AI exposure documented by Del Rio-Chanona, Laurentsyeva, and Wachs (2024) and the heterogeneous effects across knowledge communities documented by Quinn and Gutt (2025). A pre-AI proxy for π that does not reference AI capability is the platform's answer reuse rate: the fraction of answers that are viewed or linked from other questions at high frequency, indicating that the underlying problems are common and procedural rather than novel. Cross-platform variation in this proxy should predict differential exposure to the AI-induced decline, though the proxy measures a consequence of the routine-task share rather than the share itself, and more direct alternatives (such as expert classification of a question sample) may be preferable where available.

5 Public Logging and AI-to-Commons Conversion

The previous sections established that AI can reduce archive creation through two channels: the flow margin (fewer knowledge-enhancing queries reach the platform) and the resolution margin (contributor pool thinning lowers the probability that posted queries are resolved). A natural governance question is whether platform design can mitigate one or both channels. We study a conversion lever that targets the flow margin directly: public logging that converts private AI-assisted resolutions into public knowledge.

The conversion rate $\eta \in [0, 1]$ in (13) captures this channel: when private resolution succeeds in the AI regime for a type-H query, the solution enters the public archive with probability η. The reduced form accommodates a range of implementable mechanisms: default sharing of AI-assisted solutions, structured archiving that turns chat transcripts into searchable

documentation, "post your solved query" prompts, institutional logging requirements, and platform rules that encourage transforming private assistance into verifiable public artifacts. In practice, η is the product of at least two components: detection (does the system recognize the resolution as type H and potentially reusable?) and conversion (does the user actually share it?). A natural implementation is LLM-platform collaboration: when an LLM searches a platform such as Stack Overflow in the course of assisting a user and finds no existing answer, this is a signal that the resolution may be novel and reusable. The LLM can then nudge the user to share their solution (e.g., "no existing answer covers this; would you like to post yours?"). The friction is low because the artifact already exists in the chat.

5.1 Conversion and the average creation rate

Under conversion rate η, the average creation rate in the AI economy shifts to

$$\phi^{\mathrm{AI}}(K;\eta) = \phi^{\mathrm{AI}}(K;0) + \Delta(1-\pi)\eta\,\frac{a_H^{\mathrm{AI}}(K)}{K}. \tag{23}$$

The added term is the per-unit-of-K contribution of private AI resolutions that are logged. It is bounded below by a positive constant for sufficiently small K (because $a_H^{\mathrm{AI}}(K)/K$ is bounded away from zero near the origin; see the assumption below), and it vanishes as $K \to \infty$ because a_H^{AI} is bounded while K grows. This asymmetry is what gives conversion its leverage at the low-archive margin: precisely where the platform is most vulnerable to the self-undermining feedback of Proposition 3, the conversion term provides a non-vanishing boost to average creation.

Proposition 4 (Conversion and expansion of the viable region)**.** *Maintain the conditions of Proposition 3 for $\phi^{\mathrm{AI}}(K;0)$. Suppose $a_H^{\mathrm{AI}}(K)$ is continuous on $(0,\infty)$, strictly positive for all $K>0$, and satisfies*

$$\liminf_{K\downarrow 0} \frac{a_H^{\mathrm{AI}}(K)}{K} > 0. \tag{24}$$

Then:

(i) For any $\eta > 0$, $\phi^{\mathrm{AI}}(K;\eta) > \phi^{\mathrm{AI}}(K;0)$ for all $K > 0$: conversion strictly raises average creation.

(ii) *The added term $\Delta(1-\pi)\eta \cdot a_H^{\mathrm{AI}}(K)/K$ is bounded below by a positive constant for sufficiently small K and vanishes as $K \to \infty$.*

(iii) *There exists a critical conversion rate $\bar{\eta} > 0$ such that for all $\eta > \bar{\eta}$ the low-archive basin is eliminated: $\phi^{\mathrm{AI}}(K;\eta) > \lambda$ for all $K \in (0, K_U^{\mathrm{AI}}]$. If $\bar{\eta} \leq 1$, the basin can be eliminated by an implementable conversion rate. If $\bar{\eta} > 1$, even perfect conversion ($\eta = 1$) is insufficient and complementary instruments are needed to sustain the archive. The critical rate is*

$$\bar{\eta} = \sup_{K \in (0, K_U^{\mathrm{AI}}]} \frac{[\lambda - \phi^{\mathrm{AI}}(K;0)]_+}{\Delta(1-\pi)\, a_H^{\mathrm{AI}}(K)/K}. \tag{25}$$

(iv) *$\bar{\eta}$ is weakly decreasing in pointwise increases of $a_H^{\mathrm{AI}}(K)/K$ on $(0, K_U^{\mathrm{AI}}]$: the better AI is at resolving H-type queries privately when the archive is thin, the less conversion is needed to eliminate the structural threshold.*

The proof is provided in Appendix A. The condition (24) requires that $a_H^{\mathrm{AI}}(K)$ does not vanish faster than linearly at the origin; it is satisfied by any specification differentiable at zero with positive derivative, including the parametric example $a_H^{\mathrm{AI}}(K) = 1 - e^{-\rho K}$ (where the ratio $a_H^{\mathrm{AI}}(K)/K \to \rho$). The critical rate $\bar{\eta}$ in (25) is the pointwise worst-case ratio of the basin shortfall to the conversion leverage: it equals zero when $\phi^{\mathrm{AI}}(K;0) \geq \lambda$ throughout the threshold region (no basin to fill) and is large when the shortfall is deep relative to $a_H^{\mathrm{AI}}(K)/K$. Whether $\bar{\eta} \leq 1$ depends on this depth-to-leverage ratio: when the basin is shallow, moderate conversion suffices; when the self-undermining feedback is severe, even full conversion cannot offset the gap, and the platform requires resolution-side instruments to survive. This reinforces the paper's central message that the two margins require different instruments: logging alone may be insufficient when the contributor pool has collapsed.

Part (iv) contains an important implication: the conversion threshold $\bar{\eta}$ is lower in domains where AI is already good at resolving knowledge-enhancing problems privately even without a rich archive. This is precisely the case where the flow margin is largest (much of H-type creation happens off-platform), so there is a large stock of private resolutions available for logging. In other words, the domains most damaged by the flow margin are also the domains where logging is most effective at lowering the structural threshold.

5.2 Scope and limits of conversion

Conversion offsets the decline associated with private diversion by logging resolutions that would otherwise never enter the public archive. But it does not, by itself, repair congestion- or participation-driven deterioration in conditional resolution. Even at $\eta = 1$, questions that reach the platform and go unanswered because the contributor pool has thinned represent a knowledge loss that conversion cannot recover. The two components of the decline identified in Corollary 1 respond to different instruments: η targets private diversion (Channel 1), while sustaining conditional resolution requires maintaining contributor engagement through complementary instruments such as reputation systems, visibility rewards, community norms, and platform design that preserves the return to volunteering.

This scope limitation has a concrete empirical implication. If the resolution margin accounts for a substantial share of the total decline in archive creation, as preliminary evidence from Stack Overflow suggests, then conversion alone would be insufficient to restore the pre-AI creation rate. A platform that implements high η but allows its contributor pool to thin would log more private resolutions while losing the capacity to resolve the harder, knowledge-enhancing problems that reach the platform. The two instruments are complements, not substitutes.

The model does not claim that increasing η is costless or unambiguously welfare improving. Real implementations may entail privacy costs (some private queries contain sensitive information), intellectual property costs (firms may not want problem-solving strategies public), curation costs (unfiltered AI outputs can degrade the archive rather than sustain it), and compliance or legislative costs. The model isolates the archive-stock externality and identifies η as the lever that addresses the flow margin; optimal policy design that balances these costs against archive preservation is left to future work.

6 Discussion

The diffusion of AI changes how open knowledge platforms generate and sustain public archives. The private benefit is immediate, as discussed. The social risk studied here, on the other hand, is narrower and dynamic: when AI strengthens private outside options on both the user side and the contributor side, the flow of publicly logged solutions can fall enough to reduce the long-run archive stock that supports future problem-solving. This is one concrete instance of

the broader "AI dilemma" emphasized by Jones (2024): private gains can coexist with long-run risks that are hard to price.

The two-margin decomposition yields empirically identifiable predictions. Two observable patterns would help distinguish the channels identified in this paper. First, if resolution declines on a platform where the active contributor pool is stable (measured, for instance, by unique answerers per period), pool thinning is not the source; the decline must reflect other forces such as changes in moderation or closure policy, reduced effective query lifetime, or the complementary reputational-dilution mechanism documented by Li and Kim (2025) (see Remark 1). Second, if posted volume falls but the conditional resolution rate rises, the platform is in the composition-dominant regime of Proposition 1, where the J. S. Gans (2024) selection logic applies and pool thinning is absent or weak; in that case only the flow margin is active, the case for resolution-side instruments is weaker, and conversion via η may suffice on its own.

We turn to a more detailed comparison with Acemoglu, Kong, and Ozdaglar (2026), as previewed in the introduction. Our paper shares the broad concern that AI-induced private efficiency can reduce the stock of open knowledge, but the mechanisms and policy implications differ. Their framework operates through a single aggregate effort variable and does not distinguish the posted-flow margin from the resolution margin; the policy instrument that emerges from their analysis is garbling (deliberately degrading AI accuracy to incentivize human effort). Our multi-margin decomposition leads to a different instrument: public logging, which preserves full AI capability while converting private resolutions into public artifacts. The distinction matters because garbling restricts the private benefit of AI in order to sustain public knowledge, whereas logging captures a byproduct of private use without reducing its quality. Moreover, garbling is most effective when the representational structure of knowledge is fixed (so that degrading AI accuracy forces agents back to public problem-solving), but loses force when AI enables representational innovation, i.e., new ways of framing and solving problems that did not exist before; logging, by contrast, can capture novel representations as easily as standard ones. Neither framework dominates: Acemoglu, Kong, and Ozdaglar (2026) provide general-equilibrium welfare analysis that the present partial-equilibrium model cannot, while the two-margin decomposition here identifies platform-level channels that their aggregate framework does not separate. A further distinction is the congested-regime asymmetry between supply and demand shocks (Remark 1): in the present model, a simultaneous decline in posted volume and conditional resolution requires that supply-side pool thinning is quantitatively

present, providing a discriminating prediction that neither a framework with a single effort variable nor one with a fixed contributor pool generates.

The model isolates one channel through which AI can reduce open digital public goods: the platform-mediated conversion of problem-solving into a reusable public archive. It does not claim that AI reduces overall innovation or economy-wide knowledge. Proprietary R&D, firm-specific learning, and other channels through which AI may create or diffuse knowledge are outside the model's scope, as is multi-platform competition in which users and contributors substitute between platforms (or between platforms and AI) in response to relative archive quality. A multi-channel environment with both public and proprietary knowledge stocks and endogenous substitution between them would be needed for full welfare analysis, including the possibility that some knowledge lost from public platforms migrates to proprietary repositories. The conversion parameter η should be interpreted as conversion into public artifacts (documentation, answers, curated posts), not as "training use" of private interactions unless a separate mapping between those objects is explicitly defined and defended.

The model's architecture, however, abstracts beyond Q&A platforms: strip away the platform-specific language and what remains is stochastic arrival of problems at a knowledge frontier, a matching technology between problems and solvers, a public archive that accumulates from resolved problems and lowers future resolution costs, an outside option competing for solvers' attention, and AI entering by improving that outside option and diverting problems to private resolution. This architecture applies, with different incentive structures, to academic research (where the archive is the published literature and depreciation reflects methodological obsolescence), publicly funded policy analysis (where the archive is the stock of reports, briefs, and public datasets), and private R&D (where the archive is partly proprietary and returns are internalized through intellectual property).

The model's fit varies across these settings, and two dimensions of variation matter most. The first is the incentive to publicize: academic priority and tenure norms structurally attenuate the flow margin by making "posting" (publication) partially decoupled from the outside option, whereas Q&A platforms and wikis lack such structural incentives and are correspondingly more exposed. The second is the severity of the congestion externality: firms coordinate research assignments internally, attenuating the matching friction that drives the resolution margin in the present model, while volunteer-based platforms bear the full force of uncoordinated exit. Private R&D is a limiting case in which the externality structure is fundamentally different: as

the archive is partly proprietary, shrinkage would require a mechanism other than the congestion externality modeled here. On creative platforms, the architecture faces additional limitations that trace back to a deliberate modeling choice: the answering cost c is independent of the ability parameter α that governs exit. The baseline model therefore isolates the extensive-margin contribution channel: AI-induced outside-option improvements reduce participation and shrink the contributor pool, which reduces match rates when capacity is scarce (Proposition 1). It does not model differential on-platform productivity among remaining contributors, so it is silent on whether contributor exit also thins expertise or lowers answer quality conditional on participation. Kim, Jin, and Lee (2026) document reduced human creator activity on Pixiv following generative AI adoption, with the largest effects among the most productive incumbents; Li and Kim (2025) find that high-activity answerers on Stack Overflow disproportionately reduce contributions after ChatGPT's release. Both patterns are consistent with the supply-side pool-thinning mechanism in the present model, but they may also reflect quality-composition effects that the baseline does not capture. Extending the model to allow ability-dependent on-platform productivity would be needed to speak to quality thinning, whether on creative platforms or on Q&A platforms where expert answerers are disproportionately valuable.

Contributors who exit public platforms, moreover, do not disappear: they may produce knowledge in proprietary or academic channels, so that what the platform loses need not represent a net reduction in economy-wide knowledge creation. Whether AI-induced shrinkage aggregates across channels into something approaching the knowledge decline of Acemoglu, Kong, and Ozdaglar (2026) therefore depends on whether the channels most exposed to both margins (those with weak publication incentives and strong congestion externalities) are also the channels that matter most for cumulative knowledge creation, and on whether diverted effort substitutes into channels that produce equivalent public value. The model identifies the two margins that any cross-channel empirical investigation would need to measure, but it does not predict the sign of the aggregate; the cross-channel implications are a question for future work, not a conclusion of this paper.

References

Abis, Simona, and Laura Veldkamp. 2023. "The Changing Economics of Knowledge Production." Edited by Itay Goldstein. *The Review of Financial Studies* 37, no. 1 (December): 89–118. https://doi.org/10.1093/rfs/hhad059.

Acemoglu, Daron, Dingwen Kong, and Asuman Ozdaglar. 2026. *AI, Human Cognition and Knowledge Collapse.* Working Paper, 34910, February. https://doi.org/10.3386/w34910.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. "Generative AI at Work." *The Quarterly Journal of Economics* 140, no. 2 (April): 889–942. https://doi.org/10.1093/qje/qjae044.

Del Rio-Chanona, R Maria, Nadzeya Laurentsyeva, and Johannes Wachs. 2024. "Large Language Models Reduce Public Knowledge Sharing on Online Q&A Platforms." *PNAS Nexus* 3, no. 9 (September): pgae400. https://doi.org/10.1093/pnasnexus/pgae400.

Dillon, Eleanor, Sonia Jaffe, Nicole Immorlica, and Christopher Stanton. 2025. *Shifting Work Patterns with Generative AI.* Technical report w33795. Cambridge, MA: National Bureau of Economic Research, May. https://doi.org/10.3386/w33795.

Gans, Joshua. 2026. *A Model of Artificial Jagged Intelligence.* Technical report w34712. Cambridge, MA: National Bureau of Economic Research, January. https://doi.org/10.3386/w34712.

Gans, Joshua S. 2024. "Will User-Contributed AI Training Data Eat Its Own Tail?" *Economics Letters* 242 (September): 111868. https://doi.org/10.1016/j.econlet.2024.111868.

Jones, Charles I. 2024. "The AI Dilemma: Growth versus Existential Risk." *American Economic Review: Insights* 6, no. 4 (December): 575–590. https://doi.org/10.1257/aeri.20230570.

Kim, Sueyoul, Ginger Zhe Jin, and Eungik Lee. 2026. *Does Generative AI Crowd Out Human Creators? Evidence from Pixiv.* Technical report w34733. Cambridge, MA: National Bureau of Economic Research, January. https://doi.org/10.3386/w34733.

Kummer, Michael, Olga Slivko, and Xiaoquan (Michael) Zhang. 2020. "Unemployment and Digital Public Goods Contribution." *Information Systems Research* 31, no. 3 (September): 801–819. https://doi.org/10.1287/isre.2019.0916.

Li, Xinyu, and Keongtae Kim. 2025. “Impacts of Generative AI on User Contributions: Evidence from a Coding Q &A Platform.” *Marketing Letters* 36, no. 3 (September): 577–591. https://doi.org/10.1007/s11002-024-09747-1.

Lyu, Liang, James Siderius, Hannah Li, Daron Acemoglu, Daniel Huttenlocher, and Asuman Ozdaglar. 2025. *Wikipedia Contributions in the Wake of ChatGPT*. https://doi.org/10.48550/ARXIV.2503.00757.

Padilla, Nicolas, H. Tai Lam, Anja Lambrecht, and Brett Hollenbeck. 2025. *The Impact of LLM Adoption on Online User Behavior * https://doi.org/10.2139/ssrn.5393256.

Quinn, Martin, and Dominik Gutt. 2025. “Heterogeneous Effects of Generative Artificial Intelligence (GenAI) on Knowledge Seeking in Online Communities.” *Journal of Management Information Systems* 42, no. 2 (April): 370–399. https://doi.org/10.1080/07421222.2025.2487313.

A Proofs

A.1 Proof of Lemma 1

The expected knowledge increment is

$$\bar{\Delta}^e(K) = \frac{(1-\pi)\, q_H^e(K) \cdot \Delta}{\pi\, q_L^e(K) + (1-\pi)\, q_H^e(K)} = \Delta \cdot \frac{(1-\pi)\, q_H^e(K)}{Q^e(K)},$$

which equals Δ times the H-share of the posted pool. Writing the H-share as $(1-\pi)q_H^e/(\pi q_L^e + (1-\pi)q_H^e)$, this is increasing in q_H^e/q_L^e (immediate from the derivative of the fraction with respect to the ratio). Since $c^{*,e}(K) = \max\{0, \beta\bar{\Delta}^e(K) + u - C(K)\}$ is weakly increasing in $\bar{\Delta}^e$, we have $c^{*,\mathrm{AI}}(K) \geq c^{*,\mathrm{HO}}(K)$ iff $\bar{\Delta}^{\mathrm{AI}}(K) \geq \bar{\Delta}^{\mathrm{HO}}(K)$, iff the posted pool is weakly more H-rich under AI.

For the sufficient condition, hold escalation probabilities fixed: $m_\theta^{\mathrm{AI}} = m_\theta^{\mathrm{HO}} \equiv m_\theta$ for both θ. Then $q_\theta^e = m_\theta \cdot (1 - a_\theta^e)$, so

$$\frac{q_H^e}{q_L^e} = \frac{m_H}{m_L} \cdot \frac{1-a_H^e}{1-a_L^e}.$$

With m_H/m_L held fixed across environments, the ratio q_H^e/q_L^e rises under AI iff

$$\frac{1-a_H^{\mathrm{AI}}}{1-a_L^{\mathrm{AI}}} > \frac{1-a_H^{\mathrm{HO}}}{1-a_L^{\mathrm{HO}}},$$

which rearranges to (17). □

A.2 Proof of Proposition 1

Under congestion ($\Psi(\alpha^{*,e}) < Q^e$) in both environments and $T = 1$, the lifetime resolution probability is $\sigma^e(K) = (\Psi(\alpha^{*,e}(K))/Q^e(K)) \cdot F(c^{*,e}(K))$. The ratio of resolution probabilities factors as

$$\frac{\sigma^{\mathrm{AI}}(K)}{\sigma^{\mathrm{HO}}(K)} = \underbrace{\frac{\Psi(\alpha^{*,\mathrm{AI}}(K))}{\Psi(\alpha^{*,\mathrm{HO}}(K))}}_{\text{pool ratio}} \cdot \underbrace{\frac{Q^{\mathrm{HO}}(K)}{Q^{\mathrm{AI}}(K)}}_{\text{congestion relief}} \cdot \underbrace{\frac{F(c^{*,\mathrm{AI}}(K))}{F(c^{*,\mathrm{HO}}(K))}}_{\text{composition}}.$$

The first factor is less than one because AI raises outside options ($w^{\mathrm{AI}} > w^{\mathrm{HO}}$), which lowers $\alpha^{*,e}$ and hence $\Psi(\alpha^{*,e})$. The second factor is weakly greater than one because AI improves private resolution ($a_\theta^{\mathrm{AI}} \geq a_\theta^{\mathrm{HO}}$), which reduces posted flow Q^e; in the congested regime, lower Q

raises the match probability. The third factor is weakly greater than one under the conditions of Lemma 1 (the posted pool becomes more H-rich, raising $\bar{\Delta}^e$ and hence $c^{*,e}$ and $F(c^{*,e})$). Resolution falls ($\sigma^{\mathrm{AI}} < \sigma^{\mathrm{HO}}$) iff the product is less than one, which is equivalent to

$$\frac{\Psi(\alpha^{*,\mathrm{AI}}(K))}{\Psi(\alpha^{*,\mathrm{HO}}(K))} \cdot \frac{F(c^{*,\mathrm{AI}}(K))}{F(c^{*,\mathrm{HO}}(K))} < \frac{Q^{\mathrm{AI}}(K)}{Q^{\mathrm{HO}}(K)},$$

as stated in (19). □

A.3 Proof of Proposition 2

Maintain $\eta = 0$. At the human-only steady state, $h^{\mathrm{HO}}(K^{\mathrm{HO}}) = \lambda K^{\mathrm{HO}}$. The condition (20) gives $h^{\mathrm{AI}}(K^{\mathrm{HO}}; 0) < h^{\mathrm{HO}}(K^{\mathrm{HO}}) = \lambda K^{\mathrm{HO}}$, so

$$h^{\mathrm{AI}}(K^{\mathrm{HO}}; 0) - \lambda K^{\mathrm{HO}} < 0.$$

At $K = 0$, $h^{\mathrm{AI}}(0; 0) \geq 0$ (because $q_H^{\mathrm{AI}}(0) \geq 0$ and $\sigma^{\mathrm{AI}}(0) \geq 0$), so $h^{\mathrm{AI}}(0; 0) - \lambda \cdot 0 = h^{\mathrm{AI}}(0; 0) \geq 0$. If $h^{\mathrm{AI}}(0; 0) = 0$, then $K = 0$ is a steady state of the AI economy and the claim holds with $K^{\mathrm{AI}} = 0 < K^{\mathrm{HO}}$. If $h^{\mathrm{AI}}(0; 0) > 0$, define $\varphi(K) \equiv h^{\mathrm{AI}}(K; 0) - \lambda K$. We have $\varphi(0) > 0$ and $\varphi(K^{\mathrm{HO}}) < 0$. Under the maintained assumption that $h^{\mathrm{AI}}(\cdot\,; 0)$ is continuous, φ is continuous on $[0, K^{\mathrm{HO}}]$. By the intermediate value theorem, there exists $K^{\mathrm{AI}} \in (0, K^{\mathrm{HO}})$ such that $\varphi(K^{\mathrm{AI}}) = 0$, i.e., $h^{\mathrm{AI}}(K^{\mathrm{AI}}; 0) = \lambda K^{\mathrm{AI}}$. This is a steady state of the AI economy with $K^{\mathrm{AI}} < K^{\mathrm{HO}}$. □

A.4 Proof of Corollary 1

Under $\eta = 0$, $h^e(K; 0) = \Delta(1 - \pi) q_H^e(K) \sigma^e(K)$ for both $e \in \{\mathrm{HO}, \mathrm{AI}\}$. Therefore $h^{\mathrm{AI}}(K^{\mathrm{HO}}; 0) < h^{\mathrm{HO}}(K^{\mathrm{HO}})$ is equivalent to $q_H^{\mathrm{AI}}(K^{\mathrm{HO}}) \sigma^{\mathrm{AI}}(K^{\mathrm{HO}}) < q_H^{\mathrm{HO}}(K^{\mathrm{HO}}) \sigma^{\mathrm{HO}}(K^{\mathrm{HO}})$. For the decomposition, write $\Delta q_H \equiv q_H^{\mathrm{HO}} - q_H^{\mathrm{AI}}$, $\Delta\sigma \equiv \sigma^{\mathrm{HO}} - \sigma^{\mathrm{AI}}$, $\bar{q}_H \equiv (q_H^{\mathrm{HO}} + q_H^{\mathrm{AI}})/2$, and $\bar{\sigma} \equiv (\sigma^{\mathrm{HO}} + \sigma^{\mathrm{AI}})/2$. Expanding:

$$\begin{aligned}
\bar{\sigma}\,\Delta q_H + \bar{q}_H\,\Delta\sigma &= \tfrac{1}{2}(\sigma^{\mathrm{HO}} + \sigma^{\mathrm{AI}})(q_H^{\mathrm{HO}} - q_H^{\mathrm{AI}}) + \tfrac{1}{2}(q_H^{\mathrm{HO}} + q_H^{\mathrm{AI}})(\sigma^{\mathrm{HO}} - \sigma^{\mathrm{AI}}) \\
&= \tfrac{1}{2}(q_H^{\mathrm{HO}}\sigma^{\mathrm{HO}} - q_H^{\mathrm{AI}}\sigma^{\mathrm{HO}} + q_H^{\mathrm{HO}}\sigma^{\mathrm{AI}} - q_H^{\mathrm{AI}}\sigma^{\mathrm{AI}}) \\
&\quad + \tfrac{1}{2}(q_H^{\mathrm{HO}}\sigma^{\mathrm{HO}} - q_H^{\mathrm{HO}}\sigma^{\mathrm{AI}} + q_H^{\mathrm{AI}}\sigma^{\mathrm{HO}} - q_H^{\mathrm{AI}}\sigma^{\mathrm{AI}}) \\
&= q_H^{\mathrm{HO}}\sigma^{\mathrm{HO}} - q_H^{\mathrm{AI}}\sigma^{\mathrm{AI}},
\end{aligned}$$

where the cross terms cancel. The decomposition is an accounting identity: the flow margin $\bar{\sigma}\,\Delta q_H$ evaluates the effect of the posted-flow decline at the cross-regime average resolution rate, and the resolution margin $\bar{q}_H\,\Delta\sigma$ evaluates the effect of the resolution decline at the cross-regime average posted flow. Under midpoint weighting, these two terms exhaust the total decline with no residual interaction term. □

A.5 Proof of Proposition 3

Maintain $\eta = 0$. For $K > 0$, write $h^{\mathrm{AI}}(K;0) - \lambda K = K\big(\phi^{\mathrm{AI}}(K) - \lambda\big)$ where $\phi^{\mathrm{AI}}(K) \equiv h^{\mathrm{AI}}(K;0)/K$.

Step 1: (a) implies $\phi^{\mathrm{AI}}(K) < \lambda$ near $K = 0$. Under Assumption 1(a), $C(0) > \beta\Delta + u$. Since $\bar{\Delta}^e(K) \le \Delta$ for all K and C is continuous, there exists $\varepsilon_1 > 0$ such that $\beta\bar{\Delta}^e(K) + u - C(K) < 0$ for all $K \in [0, \varepsilon_1)$, hence $c^{*,e}(K) = 0$, $F(c^{*,e}) = 0$, $\sigma^e = 0$, $U_\theta^e = 0$, $m_\theta^e = \Gamma_\theta(0) = 0$ (since Γ_θ admits a density), $Q^e = 0$, and $h^{\mathrm{AI}}(K;0) = 0$. Therefore $\phi^{\mathrm{AI}}(K) = 0 < \lambda$ for $K \in (0, \varepsilon_1)$.

Step 2: (b) provides interior net growth. By Assumption 1(b), there exists $\hat{K} > 0$ with $\phi^{\mathrm{AI}}(\hat{K}) > \lambda$, so $h^{\mathrm{AI}}(\hat{K};0) - \lambda\hat{K} > 0$.

Step 3: (c) implies $\phi^{\mathrm{AI}}(K) < \lambda$ for large K. Under Assumption 1(c), $w^{\mathrm{AI}}(\underline{\alpha}, K; \pi) \to \infty$ as $K \to \infty$. Since $S^e(K) \le \Pi^e(K) \le \bar{c}$ is bounded, for all sufficiently large K we have $w^{\mathrm{AI}}(\underline{\alpha}, K; \pi) > S^e(K)$, so $\alpha^{*,e}(K) < \underline{\alpha}$ and $\Psi(\alpha^{*,e}) = 0$. With no contributors, $\mu^e = 0$, $\sigma^e = 0$, $h^{\mathrm{AI}} = 0$, and $\phi^{\mathrm{AI}}(K) = 0 < \lambda$.

Step 4: IVT and stability. By continuity of ϕ^{AI} on $(0, \infty)$ and the intermediate value theorem applied on $(\varepsilon_1, \hat{K})$, the set $\{K \in (\varepsilon_1, \hat{K}) : \phi^{\mathrm{AI}}(K) = \lambda\}$ is nonempty. Define $K_U^{\mathrm{AI}} \equiv \inf\{K \in (\varepsilon_1, \hat{K}) : \phi^{\mathrm{AI}}(K) = \lambda\}$. Since ϕ^{AI} is continuous, $\phi^{\mathrm{AI}}(K_U^{\mathrm{AI}}) = \lambda$, i.e., $h^{\mathrm{AI}}(K_U^{\mathrm{AI}};0) = \lambda K_U^{\mathrm{AI}}$. By Step 3, there exists $\bar{K} > \hat{K}$ with $\phi^{\mathrm{AI}}(\bar{K}) < \lambda$. Similarly, $\{K \in (\hat{K}, \bar{K}) : \phi^{\mathrm{AI}}(K) = \lambda\}$ is nonempty; define $K_H^{\mathrm{AI}} \equiv \sup\{K \in (\hat{K}, \bar{K}) : \phi^{\mathrm{AI}}(K) = \lambda\}$. By continuity, $\phi^{\mathrm{AI}}(K_H^{\mathrm{AI}}) = \lambda$.

By construction, $0 < K_U^{\mathrm{AI}} < \hat{K} < K_H^{\mathrm{AI}}$. The extremal definitions determine the sign pattern of $\phi^{\mathrm{AI}}(K) - \lambda$:

- K_U^{AI} *unstable from below.* For $K \in (\varepsilon_1, K_U^{\mathrm{AI}})$: $\phi^{\mathrm{AI}}(K) \neq \lambda$ (by infimum) and $\phi^{\mathrm{AI}}(\varepsilon_1) < \lambda$ (Step 1), so by continuity $\phi^{\mathrm{AI}}(K) < \lambda$ on this interval. Hence $h^{\mathrm{AI}}(K;0) < \lambda K$ and the

archive shrinks toward zero.

- K_U^{AI} *unstable from above.* Since $\phi^{\mathrm{AI}}(\hat{K}) > \lambda$ and $K_U^{\mathrm{AI}} < \hat{K}$, the continuous function ϕ^{AI} exceeds λ at $\hat{K}$. If $\phi^{\mathrm{AI}}(K) \leq \lambda$ for all $K \in (K_U^{\mathrm{AI}}, \hat{K})$, then $\phi^{\mathrm{AI}}(\hat{K}) \leq \lambda$ by continuity, a contradiction. So there exists $K' \in (K_U^{\mathrm{AI}}, \hat{K})$ with $\phi^{\mathrm{AI}}(K') > \lambda$, and trajectories starting slightly above K_U^{AI} move away.

- K_H^{AI} *stable from above.* For $K \in (K_H^{\mathrm{AI}}, \bar{K})$: $\phi^{\mathrm{AI}}(K) \neq \lambda$ (by supremum) and $\phi^{\mathrm{AI}}(\bar{K}) < \lambda$ (Step 3), so by continuity $\phi^{\mathrm{AI}}(K) < \lambda$ on this interval. Hence $h^{\mathrm{AI}}(K;0) < \lambda K$ and the archive shrinks toward K_H^{AI}.

- K_H^{AI} *stable from below.* Since $\phi^{\mathrm{AI}}(\hat{K}) > \lambda$ and $K_H^{\mathrm{AI}} > \hat{K}$, by a symmetric argument there exists $K'' \in (\hat{K}, K_H^{\mathrm{AI}})$ with $\phi^{\mathrm{AI}}(K'') > \lambda$, and trajectories starting slightly below K_H^{AI} move toward it.

If $K_t < K_U^{\mathrm{AI}}$, then $\phi^{\mathrm{AI}}(K_t) < \lambda$ (by the first bullet), so $K_{t+1} = (1-\lambda)K_t + h^{\mathrm{AI}}(K_t;0) < K_t$: the archive declines. By induction, the trajectory is monotonically decreasing and converges to zero or a lower steady state. □

A.6 Proof of Proposition 4

Under $\eta > 0$, average creation in the AI economy is

$$\phi^{\mathrm{AI}}(K;\eta) = \phi^{\mathrm{AI}}(K;0) + \Delta(1-\pi)\eta\,\frac{a_H^{\mathrm{AI}}(K)}{K}.$$

Part (i). For any $\eta > 0$ and any $K > 0$, the added term $\Delta(1-\pi)\eta\, a_H^{\mathrm{AI}}(K)/K > 0$ because $a_H^{\mathrm{AI}}(K) > 0$ by assumption. Hence $\phi^{\mathrm{AI}}(K;\eta) > \phi^{\mathrm{AI}}(K;0)$.

Part (ii). By (24), there exist $\varepsilon > 0$ and $\ell > 0$ such that $a_H^{\mathrm{AI}}(K)/K \geq \ell$ for all $K \in (0, \varepsilon]$. Hence the added term satisfies $\Delta(1-\pi)\eta\, a_H^{\mathrm{AI}}(K)/K \geq \Delta(1-\pi)\eta\,\ell > 0$ for all $K \in (0, \varepsilon]$: it is bounded below by a positive constant for sufficiently small K. As $K \to \infty$, $a_H^{\mathrm{AI}}(K)/K \to 0$ because the numerator is bounded (values in $[0,1]$), so the added term vanishes.

Part (iii). By Proposition 3, under $\eta = 0$ there exists an unstable threshold K_U^{AI} with $\phi^{\mathrm{AI}}(K_U^{\mathrm{AI}};0) = \lambda$, and $\phi^{\mathrm{AI}}(K;0) < \lambda$ for all $K \in (0, K_U^{\mathrm{AI}})$ (by the proof of Proposition 3, Step 4).

Eliminating the low-archive basin requires $\phi^{\mathrm{AI}}(K;\eta) > \lambda$ for all $K \in (0, K_U^{\mathrm{AI}}]$. Since

$$\phi^{\mathrm{AI}}(K;\eta) > \lambda \quad \Longleftrightarrow \quad \eta > \frac{[\lambda - \phi^{\mathrm{AI}}(K;0)]_+}{\Delta(1-\pi)\, a_H^{\mathrm{AI}}(K)/K},$$

the basin is eliminated for all K simultaneously iff $\eta > \bar{\eta}$, where

$$\bar{\eta} \equiv \sup_{K \in (0, K_U^{\mathrm{AI}}]} \frac{[\lambda - \phi^{\mathrm{AI}}(K;0)]_+}{\Delta(1-\pi)\, a_H^{\mathrm{AI}}(K)/K}.$$

The supremum is well-defined and finite because the numerator is bounded (it is at most λ, since $\phi^{\mathrm{AI}} \geq 0$) and the denominator is bounded away from zero on $(0, K_U^{\mathrm{AI}}]$: on $(0, \varepsilon]$ by Part (ii), and on $[\varepsilon, K_U^{\mathrm{AI}}]$ because $a_H^{\mathrm{AI}}(K) > 0$ and $K \geq \varepsilon > 0$ on a compact set. Moreover $\bar{\eta} > 0$ because $\phi^{\mathrm{AI}}(K;0) < \lambda$ for some $K \in (0, K_U^{\mathrm{AI}})$, so the numerator is strictly positive at that K.

If $\bar{\eta} \leq 1$, the basin can be eliminated by setting $\eta > \bar{\eta}$. If $\bar{\eta} > 1$, even $\eta = 1$ leaves some $K \in (0, K_U^{\mathrm{AI}}]$ with $\phi^{\mathrm{AI}}(K;1) \leq \lambda$, and complementary instruments addressing the resolution margin are necessary.

Part (iv). The critical rate $\bar{\eta}$ is defined as a supremum of ratios in which $a_H^{\mathrm{AI}}(K)/K$ appears in the denominator. A pointwise increase in $a_H^{\mathrm{AI}}(K)/K$ on $(0, K_U^{\mathrm{AI}}]$ weakly reduces each ratio and therefore weakly reduces the supremum. Hence $\bar{\eta}$ is weakly decreasing in pointwise increases of $a_H^{\mathrm{AI}}(K)/K$ on the threshold region. □

B General Query Lifetime $T > 1$

The baseline model sets $T = 1$: posted queries are either resolved within the period they are posted or they expire. This appendix develops the general case $T > 1$, in which posted queries remain active for T periods and expire if unresolved after T periods due to loss of visibility, obsolescence, or platform closure rules. The qualitative results carry through; what changes is that the stock of active queries exceeds the per-period inflow, congestion is correspondingly more severe, and the equilibrium match probability is determined by a scalar fixed point rather than a closed-form expression.

B.1 Per-period hazard and lifetime resolution

Under general T, the platform treats each match as a one-time opportunity: the matched agent makes a one-shot answering decision (the platform does not rematch a query to an agent who has previously declined it), so the answering cutoff remains $c^{*,e}(K) = \max\{0,\, \beta\bar{\Delta}^e(K) + u - C(K)\}$ as in the baseline. For a given match probability μ, the per-period resolution hazard for an active query is

$$\rho^e(K) = \mu^e(K) \cdot F\big(c^{*,e}(K)\big), \tag{26}$$

and the lifetime resolution probability over the T-period window is

$$\sigma^e(K) = 1 - \big(1 - \rho^e(K)\big)^T. \tag{27}$$

At $T = 1$, $\sigma^e = \rho^e$ and the general model collapses to the baseline.

B.2 Steady-state stock and the congestion fixed point

In steady state with constant inflow Q^e and constant hazard ρ, the stock of active queries is the sum of surviving cohorts:

$$X^e(K) = Q^e(K) \cdot \sum_{j=0}^{T-1} \big(1 - \rho^e(K)\big)^j = Q^e(K) \cdot \frac{1 - \big(1 - \rho^e(K)\big)^T}{\rho^e(K)}, \tag{28}$$

with the natural limit $X^e(K) = T \cdot Q^e(K)$ when $\rho^e(K) = 0$ (no queries are ever resolved, so all T cohorts survive). However, ρ^e depends on μ^e, and μ^e depends on X^e:

$$\mu^e(K) = \min\left\{1,\ \frac{\Psi(\alpha^{*,e}(K))}{X^e(K)}\right\}, \qquad \rho^e(K) = \mu^e(K) \cdot F\big(c^{*,e}(K)\big). \tag{29}$$

The stock X^e is therefore not computed directly from (28); it is determined jointly with μ^e as a scalar fixed point. Define

$$X(\mu) \equiv Q^e \cdot \sum_{j=0}^{T-1} \big(1 - \mu\, F(c^{*,e})\big)^j.$$

The fixed-point condition is

$$\mu = \min\left\{1,\ \frac{\Psi(\alpha^{*,e})}{X(\mu)}\right\}, \tag{30}$$

which pins down μ, ρ, X, and σ simultaneously. Since $X(\mu)$ is decreasing in μ (higher hazard resolves queries faster, reducing the stock), the right-hand side of (30) is weakly increasing in μ in the congested regime ($\Psi < X$), which supports existence of a fixed point under standard conditions.

B.3 Cohort identity

In steady state with constant inflow and constant hazard, the per-period number of resolved queries equals the stock of active queries times the per-period hazard:

$$X^e(K) \cdot \rho^e(K) = Q^e(K) \cdot \sigma^e(K). \tag{31}$$

This identity follows from substituting (28) and (27): $X\rho = Q \cdot \frac{1-(1-\rho)^T}{\rho} \cdot \rho = Q \cdot (1-(1-\rho)^T) = Q\sigma$. The identity is a steady-state accounting relationship, not a general flow-to-stock equation; it holds when ρ is constant over query lifetimes and inflow is constant.

The cohort identity implies that stationary knowledge creation can be written as

$$h^{\text{AI}}(K;\eta) = \Delta(1-\pi)\Big[q_H^{\text{AI}}(K) \cdot \sigma^{\text{AI}}(K) + \eta \cdot a_H^{\text{AI}}(K)\Big], \tag{32}$$

which has the same form as the baseline (13). The equality $q_H \cdot \sigma = X_H \cdot \rho$ ensures that the decomposition $h = \Delta(1-\pi) \cdot q_H \cdot \sigma +$ conversion carries through to the general case. This extension is therefore a stationary comparative-statics exercise: it characterizes how long-run outcomes vary with parameters at a fixed K, under the assumption that K does not change appreciably over a query's T-period lifetime. The baseline dynamic model ($T = 1$) does not require this assumption because the stock equals the flow and no cohort accounting is needed.

B.4 Congestion amplification

The stock X^e exceeds the flow Q^e whenever $\rho < 1$ (some queries persist beyond their first period), so the congestion ratio Ψ/X is more severe than Ψ/Q. Intuitively, a backlog of unresolved queries from previous cohorts competes with new arrivals for the same contributor pool. This captures the empirical regularity that platforms with large stocks of unanswered questions face worse congestion even at moderate inflow rates.

The qualitative comparative statics from the baseline carry through. Pool thinning (lower $\alpha^{*,e}$, hence lower $\Psi(\alpha^{*,e})$) lowers μ^e (now computed against the stock X^e rather than the flow Q^e), which lowers ρ^e and therefore σ^e. The composition channel (Lemma 1) still operates through $c^{*,e}$, which enters ρ^e via $F(c^{*,e})$. The π-vulnerability prediction (Section 4.3) is unchanged: both the flow margin and the resolution margin scale with π through the same mechanisms as in the baseline.

The main quantitative difference is that congestion is amplified: for a given inflow Q^e and contributor pool $\Psi(\alpha^{*,e})$, the match probability μ^e is weakly lower under $T > 1$ than under $T = 1$ because the stock $X^e \geq Q^e$. This amplification reinforces pool thinning: any reduction in Ψ has a larger effect on μ when the denominator is X rather than Q. At $T = 1$, $X^e = Q^e$, $\sigma^e = \rho^e$, and all expressions collapse to the baseline.

C Modeling Remarks and Extensions

C.1 Inner-loop multiplicity and uniqueness conditions

The period equilibrium defined in Section 2.7 involves a nested fixed point. The inner loop, given a participation cutoff α^* (and hence a contributor pool $\Psi(\alpha^*)$), solves for posted flows (q_L, q_H). These depend on σ^e, which depends on $\mu^e = \min\{1, \Psi(\alpha^*)/Q^e\}$ and on $c^{*,e} = \max\{0, \beta\bar{\Delta}^e + u - C(K)\}$, both of which depend on Q^e through $\bar{\Delta}^e$ and the congestion ratio. This subsection characterizes when the inner loop admits multiple solutions and provides sufficient conditions for uniqueness.

Mechanism. The self-reinforcing loop works as follows. Suppose agents expect the posted pool to be H-rich, so that $\bar{\Delta}^e$ is high. Then $c^{*,e}$ is high (agents are willing to incur higher answering costs because the expected archive benefit is large), which raises σ^e (more matches result in answers), which raises the posting benefit $U^e_\theta = \sigma^e \cdot V_\theta$ for both types. If $V_H > V_L$, the benefit increase is larger for H-type queries; if additionally the posting-cost distribution Γ_H is sufficiently elastic near U^e_H, the additional posting is disproportionately H-type, which raises the H-share of the posted pool and confirms the high $\bar{\Delta}^e$. This can generate two inner-loop fixed points for the same α^*: a "high-quality, high-response" equilibrium in which the posted pool is H-rich, the answering cutoff is high, and resolution is high; and a "low-quality, low-response"

equilibrium in which the posted pool is L-rich, the answering cutoff is low, and resolution is low.

Formal condition. Write the inner-loop mapping as $\Phi : (q_L, q_H) \mapsto (\tilde{q}_L, \tilde{q}_H)$, where $\tilde{q}_\theta = m^e_\theta(K) \cdot [1 - a^e_\theta(K)]$ is computed from the σ^e, $\bar{\Delta}^e$, $c^{*,e}$, and μ^e implied by (q_L, q_H). A fixed point satisfies $\Phi(q_L, q_H) = (q_L, q_H)$. Multiplicity arises when Φ has multiple fixed points, which requires the feedback from $\bar{\Delta}^e$ through $c^{*,e}$ and σ^e back to (q_L, q_H) to be strong enough to sustain two self-consistent compositions.

Sufficient conditions for uniqueness. Uniqueness of the inner-loop fixed point is guaranteed when the feedback is weak. Two conditions suffice jointly:

(i) *Low sensitivity of the answering cutoff to composition.* If β is small (agents care little about archive expansion) or if $\bar{\Delta}^e$ varies little across compositions (which occurs when π is near zero or one, so the posted pool is almost purely H or purely L regardless of AI), the answering cutoff $c^{*,e}$ is approximately constant across inner-loop fixed points, breaking the self-reinforcing loop.

(ii) *Low elasticity of posting to resolution.* If Γ_θ has low density near U^e_θ for both θ, changes in σ^e have little effect on escalation probabilities m^e_θ, so the composition of posted flows responds weakly to changes in the answering cutoff. In particular, if $V_H - V_L$ is small, the differential posting response that drives the composition shift is attenuated.

Under these conditions, Φ is a contraction and the inner loop has a unique fixed point. The main text does not impose these conditions; where multiplicity is possible, the results select the Pareto-dominant fixed point (the "high-quality, high-response" equilibrium).

C.2 Type-independence of answering costs

The baseline model draws answering costs from a single distribution $c \sim F$ on $[0, \bar{c}]$, independent of the query's type $\theta \in \{L, H\}$. This subsection discusses the interpretation and an extension with type-indexed costs.

The type-independence assumption does not claim that H-type queries are equally easy to answer in absolute terms. It claims that conditional on a random match between agent and query, the match-quality distribution does not depend on the L/H classification. The platform

assigns queries to agents without conditioning on type (agents do not observe type at the time of matching), so the relevant heterogeneity in public resolution is idiosyncratic match quality, not problem difficulty. Type-dependent hardness is already captured by residual difficulty r on the private-resolution side: H-type queries have a difficulty distribution G_H that governs whether they can be resolved privately, and only those that survive private resolution reach the platform. Conditional on reaching the platform, the match-quality draw c reflects how well the specific agent's knowledge fits the specific question, which is plausibly orthogonal to the coarse L/H classification. A niche question may have high r (AI cannot handle it) but draw low c for an agent who happens to have the right expertise.

Allowing type-dependent on-platform cost distributions F_L and F_H (with, say, F_H first-order stochastically dominated by F_L to capture that H-type queries are harder to answer) requires specifying the information structure at the time of answering. Under the baseline assumption that matched agents do not observe query type, type-dependent costs create an inference problem: if $F_L \neq F_H$, the realized cost draw c is informative about the query's type, and the agent's decision becomes a Bayesian problem of the form

$$\text{answer iff} \quad u - C(K) - c + \beta\Delta \cdot \Pr(H \mid c, \text{posted pool}) \geq 0,$$

where $\Pr(H \mid c, \text{posted pool})$ is the posterior belief that the matched query is type H, updated from the prior (the posted-pool H-share) using the likelihood ratio $f_H(c)/f_L(c)$. The acceptance region may not be a simple threshold in c, and the equilibrium fixed-point structure becomes substantially more complex. This hidden-type extension is not pursued here.

A simpler variant instead assumes that matched agents observe the query's type before deciding whether to answer. Under this observable-type information structure, the answering cutoffs are

$$c_L^{*,e}(K) = \max\{0,\, u - C(K)\}, \qquad c_H^{*,e}(K) = \max\{0,\, \beta\Delta + u - C(K)\},$$

since the agent knows the increment is 0 for an L-type query and Δ for an H-type query. The type-dependent cost distributions F_L and F_H affect resolution probabilities through $F_L(c_L^{*,e}(K))$ and $F_H(c_H^{*,e}(K))$ but do not enter the cutoff formulas themselves. Because the agent knows the query type in this variant, the baseline object $\bar{\Delta}^e(K)$ (the expected increment from a randomly drawn posted query) no longer governs individual answering decisions. One may still define

the average knowledge increment per resolved query as a platform-level descriptive object,

$$\widetilde{\Delta}^{e}_{\text{res}}(K) = \frac{(1-\pi)\, q^{e}_{H}(K) \cdot F_{H}(c^{*,e}_{H}(K)) \cdot \Delta}{\pi\, q^{e}_{L}(K) \cdot F_{L}(c^{*,e}_{L}(K)) + (1-\pi)\, q^{e}_{H}(K) \cdot F_{H}(c^{*,e}_{H}(K))},$$

but this is an aggregate accounting identity, not the decision-relevant object for the matched agent. Since the observable-type variant changes the information structure relative to the main model and is orthogonal to the paper's principal mechanisms (posted-flow diversion and participation-driven congestion), it is not pursued in the main text.

C.3 Regime-dependent stabilizer

The congestion matching technology provides a partial stabilizer against contributor exit, but the stabilizer operates asymmetrically across the congested and uncongested regimes. This subsection formalizes the non-linearity.

In the uncongested regime ($\Psi(\alpha^{*,e}) > Q^{e}$), the match probability per query is $\mu^{e} = 1$ and the match probability per participant is $Q^{e}/\Psi(\alpha^{*,e}) < 1$. The participation surplus is $S^{e}(K) = (Q^{e}/\Psi(\alpha^{*,e})) \cdot \Pi^{e}(K)$. As agents exit (lower $\alpha^{*,e}$, lower Ψ), each remaining participant is matched more often: Q^{e}/Ψ rises, which raises $S^{e}(K)$ and partially offsets exit pressure on the participation cutoff. This is the stabilizer: moderate exit is self-limiting because it raises the return to remaining participants.

In the congested regime ($\Psi(\alpha^{*,e}) < Q^{e}$), the match probability per query is $\mu^{e} = \Psi/Q^{e} < 1$ and the match probability per participant is 1 (every participant is matched). The participation surplus is $S^{e}(K) = \Pi^{e}(K)$, which does not directly depend on Ψ (though Π^{e} may still respond to changes in Ψ through inner-loop composition effects on $\bar{\Delta}^{e}$ and $c^{*,e}$). Further thinning of the pool has no offsetting effect on participant match rates: the stabilizer is exhausted.

The transition between regimes occurs at $\Psi(\alpha^{*,e}) = Q^{e}$. At this boundary, $S^{e}(K)$ is continuous but its derivative with respect to $\alpha^{*,e}$ is not: in the uncongested regime, $\partial S^{e}/\partial \alpha^{*,e} < 0$ (exit raises S^{e}); in the congested regime, $\partial S^{e}/\partial \alpha^{*,e} = 0$ (exit has no offsetting effect through participant match rates, though Π^{e} may still change if exit affects $\bar{\Delta}^{e}$ through the inner loop). This non-linearity creates a natural fragility threshold: the platform is resilient to moderate exit while uncongested but becomes fragile once it tips into congestion. Whether the stabilizer is strong enough to prevent tipping depends on the relative magnitudes of $\partial w^{e}/\partial K$ (which drives

exit through outside-option improvement) and $\partial S^e/\partial \Psi$ (which offsets exit through congestion relief) evaluated at $\Psi \approx Q^e$.

In the context of Proposition 3, the stabilizer helps explain why the empty-archive shutdown (Assumption 1(a)) and the unbounded outside options (Assumption 1(c)) can coexist with an interior region of net growth (Assumption 1(b)). At intermediate K, the platform may be uncongested, so the stabilizer supports participation; at high K, the outside-option improvement $w_K^e > 0$ eventually pushes enough agents out to tip the platform into congestion, at which point the stabilizer is exhausted and further exit accelerates decline.

C.4 Atomistic-agent interpretation of β

The parameter $\beta \geq 0$ enters the answering payoff as the agent's valuation of the archive expansion generated by their answer: the matched agent receives $\beta \bar{\Delta}^e(K) + u$ from answering. The archive is a non-excludable public good whose expansion benefits all agents regardless of whether they participated or answered. In a continuum of agents, each individual answer has measure-zero impact on the aggregate state K: a single resolved query adds Δ to an archive of size K, which is negligible in the aggregate.

A fully rational agent who understands the continuum structure would set $\beta = 0$, since their individual contribution has no measurable effect on K and hence no effect on their own future payoffs or anyone else's. The parameter $\beta > 0$ therefore cannot be interpreted as a rational calculation of marginal impact on the aggregate state. Instead, β should be interpreted as one of the following: an internalized norm (the agent values contributing to the commons for its own sake, as in warm-glow preferences); a perceived pivotality (the agent acts as if their answer matters for the archive, perhaps because platform interfaces highlight individual contributions); a community identity motive (the agent derives utility from being a contributor to a public good, independent of aggregate impact); or a direct consumption value from a richer commons that the agent treats as partially attributable to their own contribution.

The β term in the answering payoff reflects the agent's awareness that their answer is pivotal for the specific query they are matched with (which expires if unanswered under $T = 1$, or whose resolution probability they discretely affect under general T), combined with their valuation of the resulting archive increment. The pivotality for the specific query is real (answering determines whether the query is resolved); the aggregate impact on K is negligible. The β

parameter captures the former without requiring the latter to be well-defined.

This interpretation is standard in the public-goods literature on voluntary contribution, where warm-glow or impure-altruism motives are introduced precisely because pure altruism generates a free-rider problem that would predict zero contribution in large populations. The model does not take a position on which microfoundation of β is most appropriate; the answering cutoff $c^{*,e}(K) = \max\{0,\, \beta\bar{\Delta}^e + u - C(K)\}$ is consistent with any of them.

C.5 Parametric example

The following parametric example is illustrative, not calibrated: its purpose is to verify that the conditions of Assumption 1 can be satisfied simultaneously by explicit functional forms and to exhibit the hump shape of $\phi^{\mathrm{AI}}(K)$ in a concrete instance of the model.

All primitives are specified with standard closed-form distributions. The outside-option payoff is linear in ability, archive stock, and the routine-task share:

$$w^e(\alpha, K; \pi) = \alpha\,(1 + \gamma_w^e K)\,(1 + \pi\,\delta_w^e), \tag{33}$$

$$C(K) = \frac{\bar{C}}{1 + \kappa K}, \tag{34}$$

so that $C'(K) < 0$ and w^e is increasing in α and K (strictly so when $\gamma_w^e > 0$). The answering cost is uniform, $F = \mathrm{Uniform}[0, \bar{c}]$, yielding $F(c) = \min(c, \bar{c})/\bar{c}$ and expected answering surplus $\Pi = (c^*)^2/(2\bar{c})$ when $c^* \le \bar{c}$ (which holds throughout the active region of this example). The ability distribution is uniform, $\Psi = \mathrm{Uniform}[\underline{\alpha}, \bar{\alpha}]$. Private resolution probabilities are exponential in the archive: $a_H^e(K) = 1 - e^{-\rho^e K}$ and $a_L^e(K) = 1 - e^{-\rho_L^e K}$ with $\rho_L^e > \rho^e$, so that routine queries are resolved privately at a higher rate. The posting-cost distribution is exponential, $\Gamma_\theta(x) = 1 - e^{-x/\bar{d}}$, with density on $[0, \infty)$ and mean $\bar{d}$. The cost function $C(K)$, the distributions F, Ψ, and Γ_θ, and the posting values V_H, V_L are environment-independent; only the outside-option parameters (γ_w^e, δ_w^e) and the private resolution rates (ρ^e, ρ_L^e) differ across environments.

Table 1 reports the parameter values. The maintained assumption $w^{\mathrm{AI}}(\alpha, K; \pi) \ge w^{\mathrm{HO}}(\alpha, K; \pi)$ holds for all (α, K) because $\gamma_w \ge \gamma_w^{\mathrm{HO}}$ and $\delta_w \ge \delta_w^{\mathrm{HO}}$. The period equilibrium conditions (E1)–(E8) are solved numerically on a dense grid of K values by a nested fixed-point algorithm: an inner loop iterates on the posted flows (q_L, q_H) at given participation cutoff α^* until conver-

Table 1: Parameter values for the parametric example.

Symbol	Description	Value	Symbol	Description	Value
λ	Depreciation rate	0.15	$\underline{\alpha}$	Ability lower bound	0.001
π	Routine-task share	0.4	$\bar{\alpha}$	Ability upper bound	0.2
Δ	Knowledge increment	1.0	γ_w	Outside-option K-sensitivity (AI)	0.5
β	Archive concern	0.9	δ_w	Outside-option π-sensitivity (AI)	0.5
u	Private answering benefit	0.3	γ_w^{HO}	Outside-option K-sensitivity (HO)	0.0
$\bar{C}$	Answering cost at $K=0$	1.25	δ_w^{HO}	Outside-option π-sensitivity (HO)	0.0
κ	Cost decay rate	5.0	ρ	H-type private resolution rate (AI)	0.5
$\bar{c}$	Answering cost support	1.0	ρ_L	L-type private resolution rate (AI)	1.0
V_H	Private value of H-type answer	2.0	ρ^{HO}	H-type private resolution rate (HO)	0.1
V_L	Private value of L-type answer	1.0	ρ_L^{HO}	L-type private resolution rate (HO)	0.3
$\bar{d}$	Mean posting cost	0.5			

gence, and an outer loop finds α^* by Brent's method on the residual $S^e(K) - w^e(\alpha^*, K; \pi) = 0$. Knowledge creation is $h^e(K;0) = \Delta(1-\pi)\, q_H^e(K)\, \sigma^e(K)$ at conversion rate $\eta = 0$.

Figure 1 (in Section 4.2) displays the resulting average creation rates $\phi^{\mathrm{HO}}(K) \equiv h^{\mathrm{HO}}(K)/K$ and $\phi^{\mathrm{AI}}(K) \equiv h^{\mathrm{AI}}(K;0)/K$. The three conditions of Assumption 1 are verified for the AI environment as follows. Condition (a) holds because $C(0) = \bar{C} = 1.25$ exceeds $\beta\Delta + u = 1.20$; since $\bar{\Delta}^{\mathrm{AI}}(K) \le \Delta$ for all K, the answering cutoff satisfies $c^{*,\mathrm{AI}}(K) \le \max\{0,\, \beta\Delta + u - C(K)\}$, so $c^{*,\mathrm{AI}}(0) = 0$, which implies $\sigma^{\mathrm{AI}}(0) = 0$ and $\phi^{\mathrm{AI}}(0) = 0$. Condition (b) holds because ϕ^{AI} peaks at approximately 0.49 near $K \approx 0.51$, well above $\lambda = 0.15$ (the peak-to-λ ratio is approximately 3.2), and crosses λ twice: from below at $K_U^{\mathrm{AI}} \approx 0.15$ and from above at $K_H^{\mathrm{AI}} \approx 1.55$. Condition (c) holds because $w^{\mathrm{AI}}(\underline{\alpha}, K; \pi) = \underline{\alpha}(1 + \gamma_w K)(1 + \pi\delta_w)$ grows without bound in K (since $\gamma_w > 0$), so that the participation surplus $S^{\mathrm{AI}}(K)$, which is bounded above by $\max\{(c^*)^2/(2\bar{c}),\, c^* - \bar{c}/2\} \le \beta\Delta + u$, is eventually dominated by the outside option for every agent, driving $\Psi(\alpha^{*,\mathrm{AI}}(K))$ to zero and $\phi^{\mathrm{AI}}(K)$ to zero for large K.

Under the human-only environment, ϕ^{HO} peaks at approximately 0.59 near $K \approx 0.47$ and crosses λ twice: from below at $K_U^{\mathrm{HO}} \approx 0.14$ and from above at $K^{\mathrm{HO}} \approx 2.64$, so the human-only economy also has a structural minimum viable archive of similar magnitude to the AI economy ($K_U^{\mathrm{HO}} \approx K_U^{\mathrm{AI}}$), confirming that the empty-archive shutdown is environment-independent. The stable steady state K^{HO} is roughly 70% larger than K_H^{AI}, and the viable region (K_U^{HO} to K^{HO}, width ≈ 2.50) is substantially wider than under AI (K_U^{AI} to K_H^{AI}, width ≈ 1.41). Because $\gamma_w^{\mathrm{HO}} = 0$ and the private resolution rates $\rho^{\mathrm{HO}}, \rho_L^{\mathrm{HO}}$ are lower than their AI counterparts, the outside option does not grow with the archive and queries are resolved privately at a slower rate, so the hump

is taller and wider than under AI. The comparison confirms that the self-undermining feedback introduced by AI compresses the hump and pulls K_H inward, consistent with the mechanism described in Section 4.2.

The hump shape arises from the interaction of two forces. At low K, the answering cost $C(K)$ is high, suppressing the answering cutoff and hence knowledge creation; as K rises and $C(K)$ falls, $c^{*,e}$ and σ^e increase, generating the ascending limb of the hump. At high K, private resolution rates $a^e_\theta(K)$ approach one and (under AI) the outside option w^{AI} rises, so that fewer queries are posted and fewer agents participate; the resulting decline in h^e outpaces the growth in K, producing the descending limb. The two steady states K^{AI}_U and K^{AI}_H correspond to the intersections of ϕ^{AI} with λ, as characterized by Proposition 3.

Table 2 reports a one-at-a-time sensitivity analysis around the baseline, varying γ_w (outside-option K-sensitivity), κ (answering cost decay rate), and ρ (H-type private resolution rate) over the ranges $\{0.3, 0.5, 0.7\}$, $\{3.0, 5.0, 7.0\}$, and $\{0.3, 0.5, 0.7\}$ respectively. All six variations preserve the two-crossing structure: the hump shape that generates multiple steady states under Assumption 1 is not a knife-edge of the baseline calibration. In this example, the two-crossing structure is most sensitive to κ, which governs how fast $C(K)$ falls and hence where the platform activates; it is moderately sensitive to ρ, which controls how fast private resolution erodes posted flow and which has its largest effect on the stable steady state K^{AI}_H; and it is least sensitive to γ_w, whose effect on participation operates indirectly through a longer chain of intermediate links (outside options → participation cutoff → contributor pool → match probability). Because γ_w and ρ are AI-specific parameters, the human-only steady state K^{HO} changes only when κ varies (since $C(K)$ is environment-independent). These rankings are specific to the chosen functional forms and parameter ranges; different specifications for $C(K)$, $a^e_H(K)$, or w^e could alter the relative sensitivity.

Table 2: Sensitivity of the parametric example to one-at-a-time variation of γ_w, κ, and ρ around the baseline of Table 1.

Run	Varied parameter	Value	K_U^{AI}	K_H^{AI}	Peak ϕ^{AI}	K^{HO}	Two crossings?
0	—	—	0.15	1.55	0.49	2.64	Yes
1	γ_w	0.30	0.15	1.55	0.51	2.64	Yes
2	γ_w	0.70	0.15	1.55	0.46	2.64	Yes
3	κ	3.00	0.31	1.44	0.29	2.47	Yes
4	κ	7.00	0.09	1.60	0.67	2.71	Yes
5	ρ	0.30	0.14	2.07	0.55	2.64	Yes
6	ρ	0.70	0.15	1.25	0.43	2.64	Yes